Charles University in Prague

First Faculty of Medicine

Doctoral study program:

Molecular and Cell Biology, Genetics and Virology

FIRST FACULTY
OF MEDICINE
Charles University

Ing. Šárka Šestáková

Genetic and epigenetic mechanisms (and their cooperation) in the leukemogenesis of acute myeloid leukemia in adults

Genetické a epigenetické mechanismy (a jejich kooperace) v procesu leukemogeneze akutní myeloidní leukémie dospělých

Ph.D. thesis

Supervisor: MUDr. Mgr. Cyril Šálek, Ph.D.

Advisor: Mgr. Hana Remešová, Ph.D.

Prague 2020

I hereby declare that this thesis was written solely by me and all used sources of information were properly referenced. I also declare that this work has not been submitted, wholly or substantially, either for other academic award or for a qualification at any other institution.

I agree with the permanent deposition of the electronic version of my thesis in the interuniversity database system Theses.cz in order to continuously control the details of qualification works.

In Prague 13.11. 2020

ŠÁRKA ŠESTÁKOVÁ

……

Identification entry

ŠESTÁKOVÁ, Šárka. *Genetické a epigenetické mechanismy (a jejich kooperace) v procesu leukemogeneze akutní myeloidní leukémie dospělých [Genetic and epigenetic mechanisms (and their cooperation) in the leukemogenesis of acute myeloid leukemia in adults]*. Praha 2020, 130 s., 2 přílohy. Disertační práce. Univerzita Karlova v Praze, 1. lékařská fakulta, Ústav klinické a experimentální hematologie 1. LF UK v Praze a ÚHKT. Školitel: MUDr. Mgr. Cyril Šálek, Ph.D.

I would like to thank my supervisor MUDr. Mgr. Cyril Šálek, Ph.D. for his guidance throughout my doctorate studies and for the clinical insight into my experiments. Special thanks go to my advisor Mgr. Hana Remešová, Ph.D. for all the knowledge, experience, advice, and support she gave me during my research.

Next, I would like to thank RNDr. Ingrid Hrachovinová, Ph.D. for kindly providing us the samples of AML patients that we lacked. I also would like to thank our partner institution University Hospital Brno for the mutational analyses preformed within our collaborative project and for the samples they provided for the validation cohort. I would like to thank Ing. Jiří Suttnar, CSc. and Ing. Alžběta Hlaváčková, Ph.D. for performing the mass spectrometry analyses. I would like to thank Mgr. Adam Pešek for the cytometry analyses, even though we did not succeed to prove our theory. I would like to thank my good friend Ing. Dávid Kundrát for his help and cooperation during the bioinformatic analyses and also MUDr. Jan Vydra for his advice on multivariate analyses.

I would like to express my gratitude to the entire AML group and the department of Genomics from the Institute of Hematology and Blood transfusion for their help with whatever I needed and for the enjoyable and inspiring working environment.

Last but not least, I would like to thank my family and especially my dear husband for their support during my whole studies.

Contents

ABSTRACT **4**
ABSTRAKT **5**
1 INTRODUCTION **6**
1.1 Acute myeloid leukemia 6
1.1.1 Classification 6
1.1.2 Prognostic factors 7
1.1.3 Treatment 8
1.2 Genetic aberrations in AML 9
1.2.1 Cytogenetics 11
1.2.2 Recurrently mutated genes in AML with prognostic significance 13
1.3 Epigenetic changes in AML 19
1.3.1 DNA methylation 19
1.3.2 DNA hydroxymethylation in AML 25
1.3.3 Other epigenetic changes in AML 25
1.4 Methods for genetic and DNA methylation studies 27
1.4.1 Next Generation sequencing in AML 27
1.4.2 Methods for DNA methylation analysis 28
1.4.3 Methods for DNA hydroxymethylation detection 31
2 AIMS OF THE THESIS **32**
3 MATERIAL AND METHODS **33**
3.1 DNA methylation and hydroxymethylation changes in AML patients with mutations in *DNMT3A*, *IDH1/2* or their combinations 33
3.1.1 Patients 33
3.1.2 Samples' preparation 35
3.1.3 Methylation and hydroxymethylation assessment with arrays 35
3.1.4 Bisulfite conversion 36
3.1.5 Pyrosequencing 36
3.1.6 Gene expression analyses 36
3.1.7 Statistics 37
3.1.8 Gene ontology analysis 37
3.1.9 DNA hydroxymethylation assessment via mass spectrometry 37
3.2 DNA methylation validation methods 38
3.2.1 Samples and DNA standards 38

3.2.2 Characterization of analyzed CpGs 38
3.2.3 MSRE analysis 38
3.2.4 Bisulfite conversion 38
3.2.5 Primer design 39
3.2.6 Pyrosequencing 42
3.2.7 MS-HRM analysis 42
3.2.8 Quantitative MS-PCR 42
3.3 DNA methylation sequencing panel 43
3.3.1 Patients 43
3.3.2 DNA methylation sequencing panel 43
3.3.3 Samples' preparation and targeted bisulfite sequencing 43
3.3.4 Sequencing data analysis 44
3.3.5 MethScore computation 45
3.3.6 Definitions and statistical analyses 45
3.3.7 Gene ontology analysis 45
4 RESULTS 46
4.1 DNA methylation and hydroxymethylation changes in AML patients with mutations in *DNMT3A*, *IDH1/2* or their combinations 46
4.1.1 DNA methylation profiles 46
4.1.2 DNA hydroxymethylation profiles 47
4.1.3 Validation of DNA methylation and hydroxymethylation data 50
4.1.4 Influence of co-mutations (other than *DNMT3A* and *IDH1/2*) on DNA methylation and hydroxymethylation profiles 51
4.1.5 Overall expression profiles 51
4.1.6 *CHFR* methylation levels as a potential prognostic biomarker 53
4.1.7 *GZMB* methylation levels as a potential prognostic biomarker 53
4.2 DNA methylation validation methods 56
4.2.1 MSRE analysis 56
4.2.2 Pyrosequencing 56
4.2.3 MS-HRM 58
4.2.4 Quantitative MS-PCR 59
4.2.5 Overall comparison of investigated methods 61
4.3 DNA methylation sequencing panel 63
4.3.1 Acquired sequencing data and their comparison with source literature 63
4.3.2 MethScore 67
4.3.2.1 MethScore in patients from intermediate risk group 77

4.3.3 *HOX* genes 82
5 DISCUSSION 84
5.1 DNA methylation and hydroxymethylation changes in AML patients with mutations in *DNMT3A*, *IDH1/2* or their combinations 84
5.1.1 Prognostically significant DNA methylation changes 86
5.2 DNA methylation validation methods 87
5.3 DNA methylation sequencing panel 89
5.3.1 Comparison of our sequencing data with source literature 89
5.3.2 MethScore as a new surrogate marker for AML 91
5.3.3 The role of DNA methylation changes associated with *HOX* genes 93
6 CONCLUSIONS 95
7 LIST OF ABBREVIATIONS 97
8 REFERENCES 101
9 SUPPLEMENTARY MATERIAL 114
10 APPENDICES 130
10.1 Appendix 1 130
10.2 Appendix 2 130

ABSTRACT

Acute myeloid leukemia (AML) is a hematopoietic malignancy characterized by great heterogeneity and clonal nature. In recent years, rapidly evolving next-generation sequencing methods provided a deep insight into the mutational background of AML. It was shown that ~ 44 % of AML patients harbor mutations in genes that regulate DNA methylation. So far, many researchers have tried to evaluate the prognostic significance of DNA methylation changes in AML, however, due to a great inconsistency in these studies, none of the reported markers were implemented into clinical practice.

The aim of this work was to further investigate the DNA methylation changes in AML patients with specific mutations and their prognostic effect. Next, we wanted to develop a new approach for a complex evaluation of prognostically significant DNA methylation aberrations.

In our first project, we assessed the overall DNA methylation, hydroxymethylation, and gene expression in AML patients with mutations in either *DNMT3A* or *IDH1/2* or their combinations. We discovered that each genetic aberration is connected with a distinct pattern of DNA hydroxy-/methylation changes that are not entirely reflected in altered gene expression. Patients with mutations in both genes exhibited a mixed DNA methylation profile most similar to healthy controls. Furthermore, we found a prognostically significant hypermethylation in an upstream enhancer of *GZMB* gene ($p = 0.035$). Prior to validation of the DNA hydroxy-/methylation levels measured with arrays in the first project, we compared four most common methods for DNA methylation validation: analysis with methylation specific restriction enzymes, pyrosequencing, methylation-specific high-resolution melting, and methylation-specific PCR. Pyrosequencing proved to be the most convenient method due to its single base resolution and easy implementation. Next, we focused on a comprehensive evaluation of prognostically significant DNA methylation changes using a custom sequencing panel. To assess a summarizing influence of various aberrations in DNA methylation on patients' prognosis we developed MethScore, a simply computed value that reliably stratified the patients with better and worse survival ($p < 0.001$). MethScore significance was verified in multivariate analyses and validated on an independent cohort of AML patients. We further showed that MethScore may be primarily helpful for stratifying the patients with intermediate risk.

Our research contributed to the knowledge of AML epigenetic background and the prognostic significance of DNA methylation. MethScore may serve as a new surrogate marker that can specify the prognosis of AML patients within the intermediate risk group.

ABSTRAKT

Akutní myeloidní leukémie (AML) je maligní hematopoetické onemocnění, které je vysoce heterogenní zejména díky své klonální podstatě. Rozvoj sekvenování nové generace umožnil důkladně prozkoumat mutační pozadí AML. Bylo zjištěno, že asi 44 % pacientů má mutaci v některém z genů ovlivňujících metylaci DNA. Od té doby již mnoho autorů publikovalo prognostický význam určitých změn v metylaci DNA u AML. Žádný z těchto poznatků však nebyl převeden do klinické praxe, především kvůli značné rozdílnosti jednotlivých studií.

Cílem této práce bylo jednak hlouběji prozkoumat změny v metylaci DNA u pacientů se specifickým genetickým pozadím a pokusit se nalézt jejich prognostický význam. Dále jsme chtěli vyvinout nový způsob pro komplexní zhodnocení změn v metylaci DNA, u kterých byl již význam pro prognózu AML pacientů prokázán.

V našem prvním projektu jsme zkoumali celkové DNA metylační, hydroxymetylační a expresní profily AML pacientů s mutacemi v *DNMT3A* nebo *IDH1/2* nebo v obou těchto genech. Zjistili jsme, že každá mutace je spojena s charakteristickými změnami v hydroxy-/metylaci DNA, které ovšem nejsou zcela reflektovány změnami v genové expresi. Pacienti s mutacemi v obou genech se vyznačovali směsným DNA hydroxy-/metylačním profilem, který byl nejpodobnější vzorkům zdravých dárců. Dále jsme nalezli prognosticky významnou hypermetylaci v enhanceru genu *GZMB* ($p = 0.035$). Dříve, než jsme provedli validaci dat naměřených pomocí čipů v prvním projektu, porovnali jsme čtyři nejběžněji používané metody pro tyto účely: analýzu s použitím metylačně specifických restrikčních endonukleas, pyrosekvenaci, metylačně specifickou analýzu křivek tání s vysokým rozlišením a metylačně specifickou PCR. Pyrosekvenace byla zvolena jako nejvhodnější metoda především díky svému rozlišení na úrovni jednotlivých basí a snadnému provedení. V posledním projektu jsme se soustředili na nalezení komplexního přístupu pro hodnocení prognosticky významných změn v metylaci DNA. Navrhli jsme vlastní DNA metylační sekvenační panel a pro jeho vyhodnocení vyvinuli snadno spočítatelné MethScore, které spolehlivě rozdělilo pacienty dle jejich přežití ($p < 0.001$). Význam MethScore pro hodnocení prognosy pacientů byl dále ověřen v multivariantní analýze a validován na nezávislé kohortě AML pacientů. Ukázali jsme, že MethScore jako stratifikátor prognosy velmi dobře funguje i u pacientů se středním rizikem.

Naše práce přispívá k rozšíření vědomostí o epigenetické podstatě AML a prognostickém významu metylace DNA. Námi zavedené MethScore má nesporný potenciál pro zpřesnění prognosy AML pacientů se středním rizikem.

1 INTRODUCTION

1.1 Acute myeloid leukemia

Acute myeloid leukemia (AML) is a highly heterogenous hematological malignancy. This aggressive clonal disease is characterized by poorly or abnormally differentiated cells of the myeloid hematopoietic system, called blasts, that infiltrate blood, bone marrow, and other tissues. Blast cells develop from hematopoietic progenitor cells that acquire genetic mutations as well as other aberrations (such as epigenetic). Thus, they are capable of fast proliferation and self-renewal. Their accumulation leads to impaired hematopoiesis and eventually to bone marrow failure (Döhner *et al.*, 2015; Papaemmanuil *et al.*, 2016).

In adults, AML is the most common acute leukemia, but still it is a sporadic disease (e.g. representing only 1.1 % of newly diagnosed cancers in the United States). Median age at diagnosis is 68 years. The incidence is approximately 4.3 new cases per 100,000 men and women per year. However, the incidence rises with age and is 17 per 100,000 per year for patients older than 65 years. Five-year relative survival for AML is ~ 28 % (for US population) (Doubek & Mayer, 2013; Ley *et al.*, 2013; National Cancer Institute, 2020).

Clinical symptoms of AML are usually unspecific manifestations of cytopenia such as fatigue, headache, dyspnea, hematomas, and recurrent infections. AML is diagnosed when more than 20 % of myeloid blasts are detected in bone marrow (BM). In the presence of characteristic cytogenetic abnormalities such as t(15;17), t(8;21), inv(16), and t(16;16), AML is diagnosed despite the number of blasts in BM. Myelocytic or monocytic origin of the blast cells is verified using flow cytometry and cytochemistry. Further AML classification requires thorough morphological, immunophenotyping, cytogenetic, and genetic characterization of the disease (Šálek, 2013; Arber *et al.*, 2016; Döhner *et al.*, 2017).

1.1.1 Classification

There are two major classification systems for AML (**Table 1**). The French-American-British (FAB) was established in the mid-70s and divides AML into seven subtypes according to the blast cell maturation level and the type of ancestral hematopoietic cell that the leukemia developed from (Bennet *et al.*, 1976). The WHO classification, last updated in 2016, is more thorough and takes into account also immunophenotyping criteria and the latest findings in diagnosis and prognosis of AML (Arber *et al.*, 2016).

Table 1 FAB and WHO classifications of AML (The American Cancer Society, 2018)

FAB subtype	WHO classification of AML and related neoplasms
	AML with recurrent genetic abnormalities
	AML with t(8;21)(q22;q22.1);*RUNX1-RUNX1T1*
	AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22);*CBFB-MYH11*
M3	APL with *PML-RARA*
	AML with t(9;11)(p21.3;q23.3);*MLLT3-KMT2A*
	AML with t(6;9)(p23;q34.1);*DEK-NUP214*
	AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2*, *MECOM*
	AML (megakaryoblastic) with t(1;22)(p13.3;q13.3);*RBM15-MKL1*
	AML with mutated *NPM1*
	AML with biallelic mutations of *CEBPA*
	Provisional entity: AML with BCR-ABL1, AML with mutated RUNX1
	AML with myelodysplasia-related changes
	Therapy-related myeloid neoplasms
	AML not otherwise specified (NOS)
M0	AML with minimal differentiation
M1	AML without maturation
M2	AML with maturation
M4	Acute myelomonocytic leukemia
M5	Acute monoblastic/monocytic leukemia
M6	Pure erythroid leukemia
M7	Acute megakaryoblastic leukemia
	Acute basophilic leukemia
	Acute panmyelosis with myelofibrosis
	Myeloid sarcoma
	Myeloid proliferations related to Down syndrome

1.1.2 Prognostic factors

A significant independent prognostic factor for AML is age. In older patients, there is a higher probability of adverse cytogenetic features and cumulation of genetic mutations (Tsai *et al.*, 2016). The effect of age is further modulated by general patient's health condition, performance status and other comorbidities. These factors have also a major impact to the tolerance of chemotherapy. AML type is another important prognostic factor, especially the cytogenetic and genetic characteristic of the disease. Patients with previous oncological treatment or prior myelodysplasia may develop secondary AML which has an unfavorable prognosis (Šálek, 2013; Döhner *et al.*, 2017; Leisch *et al.*, 2019).

Besides the pretreatment prognostic factors, it is also essential to monitor AML throughout the course of the disease. In particular the presence of minimal residual disease (MRD) should be assessed after treatment to evaluate the patient's response, adjust the therapy, and follow the patient's condition (Döhner *et al.*, 2017).

1.1.3 Treatment

The standard treatment of younger (< 55 years) and older patients who are physically fit is "3+7" induction chemotherapy consisting of 3 days of bolus administration of anthracycline (e.g. daunorubicin) and 7 days of continuous infusion of cytarabine. Complete remission (CR) is reached when patients attain hematologic recovery, have less than 5 % of blasts in bone marrow, and no circulating blasts. CR is achieved in 60 - 85 % of patients less than 60 years old and in 40 - 60 % of patients more than 60 years old. Achievement of CR after primary intensive treatment indicates long-term survival (Walter *et al.*, 2010). In case of failure of the intensive therapy, so called salvage protocols are employed, e.g. Fla-Ida (fludarabine, idarubicin, cytarabine) (Šálek, 2013; Döhner *et al.*, 2017).

After a patient reaches CR, consolidation therapy is used to eradicate residual leukemia. Consolidation protocols include several cycles of high-dose cytarabine (HiDAC), for patients with good risk AML, or allogeneic stem cell transplantation in case of a donor (Šálek, 2013; Doubek & Mayer, 2013).

Older patients or those with other comorbidities are unfit for intensive chemotherapy and thus receive palliative treatment with hydroxyurea, hypomethylating drugs (HMAs, e.g. azacytidine), or low dose cytarabine (Doubek & Mayer, 2013; Döhner *et al.*, 2015; Döhner *et al.*, 2017).

Recently, a number of new drugs were developed targeting specific molecules with a role in AML progression. Midostaurin, an inhibitor of FLT3 tyrosine kinase activity, was approved by FDA in combination with standard chemotherapy in 2017 for newly diagnosed *FLT3* mutation-positive AML patients. In 2018, another tyrosine kinase inhibitor, gliterinib, was approved by FDA for relapsed/refractory *FLT3* mutation-positive AML patients. And other FLT3 tyrosine kinase inhibitor, quizartinib, is currently a subject of clinical trials. Ivosidenib and enasidenib are inhibitors of isocitrate dehydrogenase 1 and 2, respectively. These drugs were approved by FDA in 2017 and 2018 as a monotherapy for patients with mutated *IDH1/2*. Venetoclax is an inhibitor of anti-apoptotic protein BCL-2 and was approved in 2018 by FDA for newly diagnosed AML patients unfit for intensive chemotherapy in combination with hypomethylating agents, decitabine and azacitidine, or low-dose cytarabine. Gentuzumab ozogamicin is an anti-CD33 antibody conjugated with cytotoxic calicheamicin which inhibits DNA synthesis. This drug was approved in 2017 by FDA for newly diagnosed CD33-positive AML in combination with standard induction therapy or as a monotherapy for relapsed/refractory CD33-positive AML (Döhner *et al.*, 2015; Leisch *et al.*, 2019; Cai & Levine, 2019).

1.2 Genetic aberrations in AML

AML is a complex and molecularly heterogenous disease. Patients that develop this malignity suffer from various somatically acquired genetic lesions (**Figure 1**). In 2013, with the rapid development of next-generation sequencing (NGS) technology and high-throughput methods, the first summarizing analysis of 200 patients was carried out examining genomic and epigenomic landscape of AML. Out of 237 mutated genes that were detected in two or more patients, 23 genes were repeatedly mutated in AML. The average number of mutations per sample was 13, quite low number in comparison with other cancers, and of these an average of 5 mutations were in genes recurrently mutated in AML (Ley *et al.*, 2013). Further investigations of AML genetic profiles revealed that patients can be divided into 11 classes with distinct phenotype and outcome according to their patterns of co-mutations (Papaemmanuil *et al.*, 2016). All the recent findings about AML genetic landscape and its prognostic significance were summarized in the 2017 recommendations of Europen LeukemiaNet (ELN), see **Table 2**. Based on these suggestions, the common clinical practice examining cytogenetic aberrations, gene-fusions, and some of the mutations (*FLT3*, *NPM1*, *CEBPA*) should be extended, at least to asses also mutations in *TP53*, *RUNX1,* and *ASXL1* (Döhner *et al.*, 2015; Döhner *et al.*, 2017). Furthermore, the integration of NGS into routine clinical examination of newly diagnosed AML patients is highly recommended (Levine & Valk, 2019).

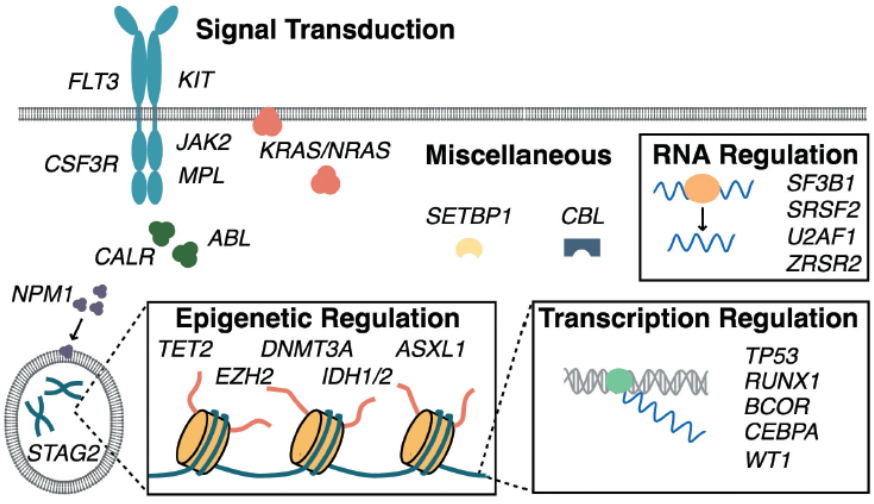


Figure 1 Genes recurrently mutated in myeloid malignancies (Tremblay *et al.*, 2018). Genes relevant to AML pathogenesis can be divided into nine groups according to their function and almost all patients have at least one mutation in a gene belonging to one of these categories: activated signaling genes (detected in 59 % of AML), DNA-methylation related genes (44 %), chromatin-modifying genes (30 %), *NPM1* mutations (27 %), myeloid transcription factor genes (22 %), transcription factor fusions (18 %), tumor suppressor genes (16 %), spliceosome-complex genes (14 %), cohesin-complex genes (13 %) (Ley *et al.*, 2013).

Table 2 Risk stratification of AML patients based on their genetic lesions (Grimwade *et al.*, 2016; Döhner *et al.*, 2017)

Risk category	Cytogenetic characteristic	Molecular-genetic characteristic
Favorable	t(8;21)(q22;q22.1)	RUNX1-RUNX1T1
	inv(16)(p13.1q22) or t(16;16)(p13.1;q22)	CBFB-MYH11
	Typically associated with normal karyotype	Mutated *NPM1* without *FLT3*-ITD^{high*} or with *FLT3*-ITD^{low*}
	Typically associated with normal karyotype	Biallelic mutated *CEBPA*
Intermediate	Typically associated with normal karyotype	Mutated *NPM1* and *FLT3*-ITD^{high*}
	Typically associated with normal karyotype	Wild-type *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low*} (without adverse-risk gene mutations)
	t(9;11)(p21.3;q23.3)	MLLT3-KMT2A
	Abnormalities not classified as favorable or adverse	
Adverse	t(6;9)(p23;q34.1)	DEK-NUP214
	t(v;11q23.3)	KMT2A rearranged
	t(9;22)(q34.1;q11.2)	BCR-ABL1
	inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2)	GATA2, MECOM(EVI1)
	−5 or del(5q); −7; −17/abn(17p)	
	Complex karyotype (≥3 abnormalities)	
	Monosomal karyotype	
		Wild-type *NPM1* and *FLT3*-ITD^{high*}
	Associated with complex and monosomal karyotype	Mutated *TP53*
		Mutated *RUNX1*†
		Mutated *ASXL1*†

*low allelic ratio (< 0.5)/high allelic ratio (> 0.5)

†Should not be used as an adverse prognostic marker if detected in favorable-risk AML subtypes

Another characteristic feature of AML genetics is clonal heterogeneity. Studies have shown that at diagnosis, there is always at least one subclone detected among the founding leukemic clone (the initial transformed cell). However, majority of mutations detected in AML are present in nearly all the cells. It is because the hematopoietic stem cells acquire these mutations before the initiating event causes the development of AML (Welch *et al.*, 2012; Ley *et al.*, 2013; Shlush *et al.*, 2014). These early mutations, called "passenger lesions", are typically detected at high variant allele frequency (VAF) in AML blasts. In most cases, these genes are involved in histone modifications, DNA methylation, and chromatin rearrangements (e.g. *DNMT3A*, *TET2*, *ASXL1*, *PPM1D, SF3B1*) and alone are not sufficient to cause AML. Mutations in these landscaping genes can also be found at low VAF in 10 – 20 % of blood samples from healthy older individuals who have thus higher risk of hematologic neoplasm development, a state called "clonal hematopoiesis of indeterminate

potential“ (Jaiswal *et al.*, 2014; Steensma *et al.*, 2015). Also, ancestral preleukemic cells with these mutations can survive chemotherapy and eventually cause relapse after expansion during remission. Nevertheless, the myeloid neoplasm develops only after acquisition of further mutations. These late, so called “driver mutations” affect usually genes from activated signaling pathways (e.g. *FLT3*, *JAK*, *NRAS*) (Corces-Zimmerman *et al.*, 2014; Döhner *et al.*, 2015; Leisch *et al.*, 2019).

1.2.1 Cytogenetics

Cytogenetic assessment of metaphase cells is an essential part of AML diagnostics since aberrant cytogenetic features have major prognostic impact and directly affect the selection of post remission therapy (Šálek, 2013; Döhner *et al.*, 2017). Abnormal karyotype is detected in about 60 % of newly diagnosed AML patients. In MRC (Medical Research Council, London) trials, Grimwade *et al.* studied the predictive value of cytogenetic changes in older patients (median age 66 years) and then revised his findings in a large meta-analysis of 5,876 younger patients (age range 16 - 59 years) (Grimwade *et al.*, 2001; Grimwade *et al.*, 2010). By these criteria, patients were divided into three major groups (**Table 3**). Acute promyelocytic leukemia (APL) and AML with translocations including core binding factor belong to the group with favorable prognosis. These patients respond well to conventional chemotherapy. On the contrary, patients within the adverse risk category, for example those with complex or monosomal karyotype, do not respond well and are preferentially indicated for allogenic transplantation or experimental treatment. Nevertheless, the majority of AML are classified as intermediate risk (IR) patients (most commonly with normal karyotype). For those, other molecular biomarkers, especially various gene mutations, are crucial for establishing the prognosis (Ley *et al.*, 2013; Šálek, 2013; Doubek & Mayer, 2013).

Table 3 Prognostic stratification based on cytogenetic changes (Grimwade *et al.*, 2010)

	Cytogenetic abnormality	
Favorable	t(15;17)(q22;q21) t(8;21)(q22;q22) inv(16)(p13q22)/t(16;16)(p13;q22)	Irrespective of additional cytogenetic abnormalities
Intermediate	Entities not classified as favorable or adverse	
Adverse	abn(3q) [excluding t(3;5)(q21~25;q31~35)], inv(3)(q21q26)/t(3;3)(q21;q26), add(5q), del(5q), −5, −7, add(7q)/del(7q), t(6;11)(q27;q23), t(10;11)(p11~13;q23), t(11q23) [excluding t(9;11)(p21~22;q23) and t(11;19)(q23;p13)] t(9;22)(q34;q11), −17/abn(17p), Complex (≥ 4 unrelated abnormalities)	Excluding cases with favorable karyotype

1.2.1.1 Fusion genes

Balanced rearrangements are detected in a significant proportion of younger AML patients (30 - 50 % in younger adults and children) and on the contrary are rare in secondary AML (Grimwade, David & Mrózek, 2011). A number of fusion genes are recognized as recurrent genetic abnormalities in AML and some of them are considered sufficient to diagnose AML irrespective of blast count in BM (Grimwade *et al.*, 2016; Arber *et al.*, 2016). The occurrence of inversion or translocation events is considered to be an initiating step in the AML development causing disruption in cells' differentiation, however, not sufficient on its own for the leukemic transformation (Miyamoto *et al.*, 2000). Fusions involving hematopoietic transcription factors (mainly *RUNX1-RUNX1T1*, *CBFB-MYH11*, *PML-RARA*) are frequently accompanied by mutations in signaling genes (e.g. *KIT*, *N/KRAS*, *FLT3*, *NF1*) as a "second hit" mutations that enhance proliferation activity of the undifferentiated cells (Jourdan *et al.*, 2013; Duployez *et al.*, 2016; Grimwade *et al.*, 2016).

PML-RARA

Acute promyelocytic leukemia (APL), a unique subtype of AML, is associated with t(15;17) leading to a formation of fusion gene PML-RARA. APL patients feature favorable prognosis and respond well to ATRA (all-*trans* retinoic acid) and ATO (arsenic trioxide) therapy (Grimwade *et al.*, 2016).

RUNX1-RUNX1T1 and CBFB-MYH11

AML with inv(16) and t(8;21), leading to *CBFB-MYH11* and *RUNX1-RUNX1T1* fusion genes, are collectively called core-binding factor (CBF) AML. Patients with CBF AML (approximately 15 % of de novo AML in adult patients) are categorized as low risk and profit from high-dose chemotherapy. Nevertheless, the relapse incidence is up to 40 % and thus the spectrum of co-mutations further defines the patients' prognosis (Šálek, 2013; Duployez *et al.*, 2016).

As mentioned previously, CBF AML rearrangements are usually associated with mutations in tyrosine kinase signaling genes. However, further investigations revealed a different spectrum of co-mutations for each subtype. *RUNX-RUNX1T1* fusions are significantly associated with mutations in cohesion complex genes (18 %) and chromatin modifiers (42 %) including *ASXL1* and *ASXL2* (35 %) (Micol *et al.*, 2014; Döhner *et al.*, 2017). On the contrary, these mutations are nearly absent in inv(16) AML. It was also shown that mutations in *KIT* and *FLT3* were associated with higher probability of relapse for CBF AML. Also, high *KIT* mutant allelic ratio has significant negative impact on t(8;21) AML prognosis and, on the contrary, high *N/KRAS* mutant allele ratios were associated with favorable outcome and a lack of *KIT* or *FLT3* mutations (Duployez *et al.*, 2016).

Other recurrent cytogenetic abnormalities

From the rest of the balanced rearrangements detected in AML patients, only *KMT2A (MLL)* is more common with incidence about 4 % of de novo AML. These translocations are in general associated with poor prognosis and rather aggressive disease. The only exception is t(9;11)/*MLLT3-KMT2A*, most common *KMT2A* rearrangement, classified as intermediate risk (Behdad & Betz, 2016).

Other cytogenetic abnormalities (*DEK-NUP214*, *GATA2, MECOM*, and *BCR-ABL1*) are less common, each type accounts only for $\sim 1-2$ % of newly diagnosed AML (Papaemmanuil *et al.*, 2016; Döhner *et al.*, 2017).

1.2.2 Recurrently mutated genes in AML with prognostic significance

NPM1

Nucleophosmin (NPM1) is a multifunctional phosphoprotein normally detected in cell nucleolus. NPM1 is involved in DNA repair, chromatin remodeling, ribosome biogenesis, replication and transcription, and have both tumor suppressive and growth promoting functions (Lindström, 2011). In 2005, mutations in the last exon (12) of *NPM1* gene were

discovered in AML (Falini *et al.*, 2005). Since then, more than forty different mutations were described with the most common mutations, detected in ~ 90 % of cases, being type A, B, and D. In the majority of cases, these mutations are four-base insertions causing disruption of nucleolar localization signal and relocation of nucleophosmin to cytoplasm.

NPM1 mutations (*NPM1*^{mut}) are the most common mutations in AML. *NPM1*^{mut} is detected in 30 % of AML and more than 50 % of cytogenetically normal AML (CN-AML) and is generally considered a founder event in AML onset. Most frequent co-mutations are detected in *DNMT3A* (50 %) and *FLT3* (40 %) genes (Falini *et al.*, 2005; Grimwade *et al.*, 2016).

NPM1^{mut} AML patients generally respond well to induction chemotherapy and have higher CR rate. *NPM1*^{mut} is thus considered as a marker of favorable prognosis. However, the overall survival (OS) and relapse ratio is markedly dependent on the co-occurrence of *FLT3*-ITD and its mutant allelic ratio (Grimwade *et al.*, 2016; Döhner *et al.*, 2017).

FLT3

FLT3 (FMS-like tyrosine kinase 3) gene encodes a hematopoiesis regulating tyrosine kinase receptor that activates downstream differentiation, proliferation and apoptosis pathways in hematopoietic stem cells via activation of RAS/RAF/MEK and PI3K/AKT cascades (Grafone *et al.*, 2012). *FLT3* is mutated in one third of all AML patients and two types of mutations can be found: internal tandem duplications (*FLT3*-ITD, detected in 27 % of AML) and point mutations in the kinase domain (*FLT3*-TKD, detected in 7 % of AML). In both cases, the gene aberration leads to constitutive activation of the receptor. However, only *FLT3*-ITD has been proved as an independent prognostic factor predicting poorer OS and higher relapse rate especially in case of higher mutant allelic burden (Grimwade *et al.*, 2016). Moreover, if *FLT3*-ITD is concurrently detected in both alleles (*FLT3*-ITD/*FLT3* wild-type ratio is higher than 0.5), the adverse effect is so strong that even in case of concurrent *NPM1*^{mut} the patients have worse prognosis and are classified as intermediate risk, see **Table 2** (Döhner *et al.*, 2017; Levine & Valk, 2019). The prognostic significance of *FLT3*-TKD mutations is less clear and is rather associated with better survival (Grimwade *et al.*, 2016).

Mutated FLT3 kinase represents an attractive therapeutic target. Therefore, many new drugs in form of small molecule tyrosine kinase inhibitors were developed and recently approved by FDA for AML treatment (e.g. midostaurin, gliterinib; see chapter 1.1.3).

CEBPA

Mutations in *CEBPA* (CCAAT/enhancer-binding protein alpha) gene, that encodes a myeloid transcription factor, are found approximately in 11 % of AML (Fasan *et al.*, 2014). The gene consists of a single exon where the mutations occur mostly in two distinct regions. Frameshift mutations are found prevalently in the *N*-terminal transactivation domain, whereas insertions and deletions in *C*-terminal domain disrupt the dimerization and DNA-binding site of the resulting protein. Interestingly, in more than half of the *CEBPA* mutated cases, both alleles are affected, each in one, either *N*-terminal or *C*-terminal, region. It was also shown, that only patients with this biallelic mutation in *CEBPA*, sometimes referred to as *CEBPA* double mutant (*CEBPA*^{dm}) AML, have favorable prognosis. Single mutations in *CEBPA* have no significant impact on patients' prognosis (Wouters *et al.*, 2009; Taskesen *et al.*, 2011).

CEBPA mutations are mutually exclusive of balanced rearrangements, usually lack *FLT3*-ITD, but are associated with *GATA2* and *NRAS* mutations (Grimwade *et al.*, 2016).

RUNX1

Gene encoding the hematopoietic transcription factor RUNX1 (runt-related transcription factor 1) is mutated in about 10 % of AML patients. Besides mutations, RUNX1 is also frequently involved in chromosomal rearrangements in AML (e.g. *RUNX1-RUNX1T1*, *MECOM(EVI1)-RUNX1*) (Gaidzik *et al.*, 2016).

RUNX1 mutations (*RUNX1*^{mut}) are mostly mutually exclusive of *FLT3*, *NPM1*, and *CEBPA* mutations and balanced rearrangements. To the co-occurring genetic lesions belong mutations in *ASXL1*, *SRSF2*, *IDH2*, and *KMT2A* gene (Ley *et al.*, 2013; Grimwade *et al.*, 2016).

Many studies confirmed that mutations in *RUNX1* indicate poor prognosis and define a group of patients with unfavorable clinicopathological features such as older age, immature morphology, or secondary AML evolved from myelodysplastic syndrome. Thus, *RUNX1*^{mut} AML was proposed as a provisional entity in 2016 WHO classification and was established as an individual category of patients with inferior outcome by the 2017 ELN recommendations (Gaidzik *et al.*, 2016; Arber *et al.*, 2016; Döhner *et al.*, 2017).

TP53

TP53 is a tumor suppressor with a key role in DNA repair, regulation of apoptosis, and cellular senescence. Mutations in *TP53* gene (*TP53*^{mut}) occur in ~ 10 % of AML patients and are mainly detected in secondary AML and older patients. *TP53*^{mut} are associated with

complex and monosomal karyotype and specific aneuploidies (e.g. deletions of chromosomes 5 or 7) (Haferlach *et al.*, 2008). *TP53*^{mut} patients have low response rate to chemotherapy and high relapse rates after stem cell transplantation. Thus, *TP53* mutations are a marker of particularly poor outcome (Rücker *et al.*, 2012; Kadia *et al.*, 2016).

1.2.2.1 Epiregulatory genes

Mutations in genes encoding epigenetic regulators, including DNA methylation and histone modifications, has been repeatedly reported by a number of groups that investigated the overall mutational landscape of AML with NGS technology (Ley *et al.*, 2013; Papaemmanuil *et al.*, 2016; Metzeler *et al.*, 2016). It is also apparent from several studies, that these mutations are not exactly AML specific, however, they define a preleukemic state that can lead to AML development through acquisition of cooperating mutations (Jaiswal *et al.*, 2014; Shlush *et al.*, 2014; Grimwade *et al.*, 2016).

DNMT3A

DNMT3A gene encodes *de novo* DNA methyltransferase. This enzyme catalyzes the addition of methyl group to the 5´-carbon of cytosine residue in a CG dinucleotide (called CpG) and thus generates a newly methylated site in the genome (Ley *et al.*, 2010).

DNMT3A mutations (*DNMT3A*^{mut}) are detected in up to 36 % of AML patients and therefore belong to the triad of most frequently mutated genes in AML together with *FLT3* and *NPM1* mutations (Wouters & Delwel, 2016). Quite commonly, these three mutations co-occur, which suggest their mutual cooperation in AML pathogenesis. It is presumed that *DNMT3A*^{mut} appear first as a founding mutation in the preleukemic clone that later acquires *NPM1* and *FLT3* mutations, see **Figure 2** (Shlush *et al.*, 2014; Corces-Zimmerman *et al.*, 2014).

Most frequent *DNMT3A*^{mut} is R882H located in the region encoding the enzyme's catalytic site. The resulting protein has a dominant-negative activity and inhibits the formation of tetramers of wild-type DNMT3A resulting in focal hypomethylation of specific CpGs throughout the genome (Russler-Germain *et al.*, 2014). Interestingly, it appears that *DNMT3A* wild-type AML possess abnormal hypermethylation in specific CpGs. The authors of the study propose that this "break" is an adaptive response of the cells that try to stop the rapid proliferation or leukemic transformation. Thus, the *DNMT3A*^{mut} patients lose this protective methylation which contributes directly to AML initiation (Spencer *et al.*, 2017).

DNMT3A^{mut} is usually seen as an adverse prognostic factor and it was demonstrated that *DNMT3A*^{mut} promotes chemoresistance (Guryanova *et al.*, 2016). However, the outcome of *DNMT3A*^{mut} AML patients depends strongly on other co-mutations, especially in case of *NPM1*^{mut} which is detected in majority (80 %) of *DNMT3A*^{mut} AML (Gale *et al.*, 2015).

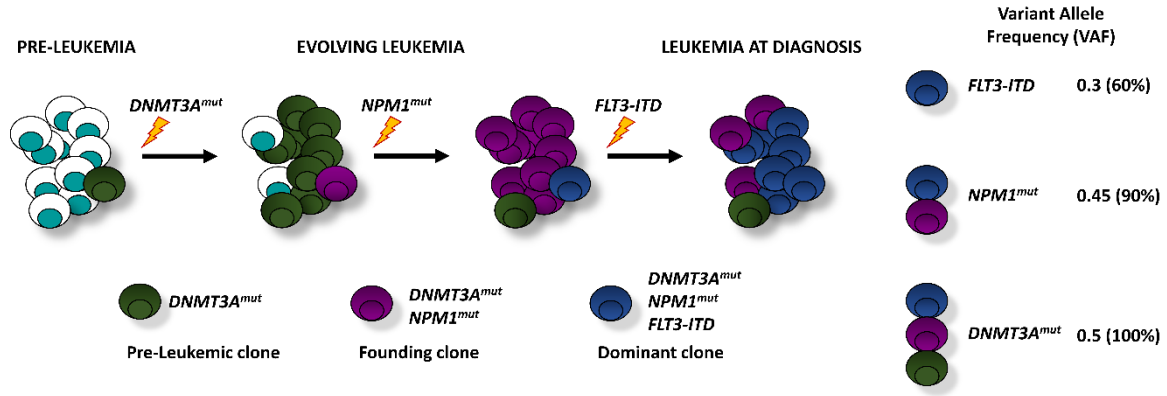


Figure 2 Hypothetical clonal evolution of AML in a patient who carries cooperating *DNMT3A*^{mut}, *NPM1*^{mut}, and *FLT3*-ITD. Mutation in *DNMT3A* appears first, prior to overt disease, and facilitates the clonal expansion. Subsequently, *NPM1*^{mut} occurs in the founding clone as a disease-defining mutation. Further acquisition of *FLT3*-ITD results in a hyperproliferative clone and leukemia expansion. This dominant clone is then detected at diagnosis. The proportion of cells carrying a specific somatic mutation in a sample can be estimated based on its allele frequency, as shown on the right (Grimwade *et al.*, 2016).

IDH1/2

IDH1 and IDH2 are cytosolic and mitochondrial isoenzymes of isocitrate dehydrogenase that catalyze one of the reactions in Krebs cycle, the oxidative decarboxylation of isocitrate to α-ketoglutarate. Mutated enzymes IDH1/2 produce an oncometabolite 2-hydroxyglutarate which inhibits the function of dioxygenase enzymes, including TET2. This leads to a global hypermethylation of the genome similar to patients with *TET2* mutations (Ward *et al.*, 2010; Montalban-Bravo & DiNardo, 2018).

IDH1/2 mutations (*IDH1/2*^{mut}) are detected in about 20 % of AML, with slightly more prevalent mutations in *IDH2* (~ 12 % of newly diagnosed patients). *IDH1*^{mut} and *IDH2* ^{mut} are mostly mutually exclusive and associated with normal karyotype. The prognostic effect of *IDH1/2* mutations presumably depends on the position of the mutation (most commonly *IDH1* R132H/C, *IDH2* R140Q and R172K), co-occurring mutations, and other genetic lesions (Papaemmanuil *et al.*, 2016; Liu & Gong, 2019). The *IDH1*^{R140} mutation is associated with *NPM1*^{mut} and indicates better prognosis. On the other hand, *IDH2*^{R172} is not detected in *NPM1*^{mut} patients and predicts a poor clinical outcome (Grimwade *et al.*, 2016). However, there is a great inconsistency across the studies targeting the prognostic significance of these

mutations (Montalban-Bravo & DiNardo, 2018). Nevertheless, new drugs with good clinical response for relapsed/refractory AML patients with mutations in *IDH1* or *IDH2* were recently approved by FDA (see chapter 1.1.3).

TET2

TET2 (ten-eleven translocation oncogene family member 2) encodes a methylcytosine dioxygenase. This enzyme catalyzes the conversion of 5′-methylcytosine (5-mC) to 5′-hydroxymethylcytosine (5-hmC), the initial step of DNA demethylation (Grimwade *et al.*, 2016).

Mutations in *TET2* are detected in ~ 8 % of patients with AML but can be detected in older healthy individuals as well, which supports the concept of pre-leukemic clonal hematopoiesis (chapter 1.2) (Jaiswal *et al.*, 2014). *TET2* mutations (*TET2*^{mut}) are mutually exclusive of *IDH1/2*^{mut} and mutations in *WT1* gene that encodes a tumor suppressor transcription factor. It was found that besides *IDH1/2*, mutations in *WT1* (found approximately in 9 % of AML) can also impair the function of TET2, in both cases leading to site-specific changes in DNA hydroxymethylation and overall genome hypermethylation (Rampal *et al.*, 2014; Wang *et al.*, 2015). These findings suggest the existence of TET2-IDH1/2-WT1 mutated AML subtype characterized by particular epigenetic changes (Grimwade *et al.*, 2016; Wouters & Delwel, 2016). Nevertheless, no clear prognostic impact of *TET2*^{mut} on the clinical outcome of AML patients was found (Gaidzik *et al.*, 2016).

ASXL1

ASXL1 (additional sex comb-like 1) belongs to the enhancers of polycomb and trithorax genes, which regulate various genes' expression via chromatin remodeling (Fisher *et al.*, 2006). Somatic mutations in *ASXL1* gene have been described in various types of myeloid malignancies including AML where it is detected in ~ 10 % of patients with higher prevalence in older patients and secondary AML (Metzeler *et al.*, 2016).

ASXL1 mutations (*ASXL1*^{mut}) are inversely associated with *FLT3*-ITD, mutually exclusive with *NPM1*^{mut}, and co-occur with *RUNX1*^{mut} and *IDH2*^{mut} (Pratcorona *et al.*, 2012; Paschka *et al.*, 2015).

Mutations in *ASXL1* have been consistently reported as an independent prognostic marker of a poor survival. Therefore, *ASXL1*^{mut} AML patients were established as a new individual category with inferior prognosis in the 2017 ELN guidelines (Döhner *et al.*, 2017).

1.2.2.2 Other genes recurrently mutated in AML

As mentioned before, with the high-throughput NGS technologies many other genes were discovered to be mutated in AML. These include signaling pathways genes, e.g. *cKIT*, *RAS*, *NF1*, *CBL*, and *PTPN11* (Grimwade *et al.*, 2016). Also, the histone modifying genes are frequently mutated in AML, especially the lysine methyltransferases (mainly *KMT2A*, also called *MLL*) and components of the polycomb repressor complexes, for example *EZH2* encoding H3K27 methyltransferase (Wouters & Delwel, 2016). Splicing factor genes are prevalently mutated in myelodysplastic syndromes and thus detected in secondary AML, however, mutations in these genes (e.g. *SRSF2*, *SF3B1*, and *U2AF1*) were also found in *de novo* AML cases (Lindsley *et al.*, 2015). In the TCGA 2013 study, the category of cohesion complex genes was identified as recurrently mutated in AML, including *RAD21*, *SMC1A*, and *STAG1* or *STAG2* genes (Ley *et al.*, 2013; Thol *et al.*, 2014).

1.3 Epigenetic changes in AML

It is not surprising that the epigenetic landscape of AML was found to be markedly disturbed, since 44 % of AML patients harbor mutations in genes encoding the regulators of DNA methylation and 43 % of AML in genes functioning as chromatin modifiers or genes of cohesion complex (Ley *et al.*, 2013). Many recent studies show that the knowledge of epigenetic background should not be neglected when establishing the AML prognosis because epigenetic changes are inseparably connected to genetic lesions and they function together as a driver mechanism in leukemogenesis (Fong *et al.*, 2014).

Furthermore, targeting the epigenome offers new therapeutic strategies, especially because epigenetic changes are mostly reversible. A variety of small molecule inhibitors are being developed affecting the epiregulatory genes. Currently approved drugs are for example the inhibitors of IDH1/2, inhibitors of histone deacetylases or DNA methyltransferases (e.g. azacytidine) (Cai & Levine, 2019).

1.3.1 DNA methylation

DNA methylation is an essential process for genome stability maintenance, proper embryonic development, gene expression regulation and cellular differentiation. In mammals, DNA methylation occurs almost exclusively on cytosines in CG dinucleotides, so called CpGs. There are over 28 milion CpGs in the human genome and most of them (~ 80 %) are methylated (Schubeler, 2015). About 10 % of CpGs are accumulated in so-called CpG islands (CGI), CpG rich regions extending for 300-3000 base pairs (bp) that are found

in nearly a half of mammalian gene promoters (Smith & Meissner, 2013). These CGIs are mostly unmethylated regardless of expression and many studies have found an association between the methylation status of a CGI and transcription of the associated gene. Moreover, aberrant hypermethylation of CGIs in promoters of tumor suppressors connected to reduced transcription of the gene and hypomethylation of non-CpG-rich regions is a well-known phenomenon in cancer (Jones, 2012). However, further studies of the epigenome revealed that changes in DNA methylation outside the CGIs, in CGI shores (regions adjacent to CpG islands up to 2 kbp away) and shelves (regions subsequent to CGI shores), and even in gene bodies may have the same or even more important role in initiation and maintenance of a malignancy (Akalin *et al.*, 2012; Schoofs *et al.*, 2014). In CGIs, the methylation usually functions as a direct obstruction for transcription factors or attracts methyl-CpG binding proteins (MBP) that recruit corepressor complexes. Mechanisms behind the gene expression regulation by differentially methylated regions (DMR) in intragenic areas are still poorly understood but probably involve regulation of elongation during transcription, determination of alternative poly-A sites and regulation of pre-mRNA splicing (Fong *et al.*, 2014).

DNA methyltransferases (DNMT) are enzymes able to add the methyl group to the fifth carbon of cytosine in CpG dinucleotide. DNMT1 maintains the methylation pattern of the genome via recognition and methylation of hemi-methylated sites, thus restoring the methylation symmetry during DNA replication. DNMT3 can methylate both hemi-methylated and unmethylated sites in the genome, helping to maintain the DNA methylation equilibrium and also being able to create newly methylated loci which is particularly important during embryogenesis and stem cell differentiation (Fong *et al.*, 2014). There are two homologs of DNMT3, DNMT3A and DNMT3B, and each preferably methylate specific sites in the genome. DNMT3B is highly expressed during early embryogenesis and then silenced in somatic cells. However, DNMT3B aberrant overexpression is tumorigenic and has been detected also in AML and connected to poor prognosis (Norvil *et al.*, 2020). Mutations in *DNMTs*, especially *DNMT3A* in AML (see chapter 1.2.2.1), are connected with overall hypomethylation of the genome (Norvil *et al.*, 2020).

Other genes affecting directly the DNA methylation are *IDH1/2* and *TET2*, described in chapter 1.2.2.1. Mutations in these genes are associated with prevailing DNA hypermethylation. Interestingly, although these mutations are mutually exclusive, the resulting hypermethylation patterns in DNA of $IDH1/2^{mut}$ and $TET2^{mut}$ patients are overlapping (Fong *et al.*, 2014).

Moreover, specific DNA methylation changes are detected in nearly all subtypes of AML with various genetic lesions. This could be partially explained by the epigenetic drift as an independent phenomenon that occurs in the course of repetitive cell cycling in aging cells or transformed cancer cells. But specific DNA methylation profiles are also connected with for example mutations in genes regulating the histone modifications, especially the histone methyltransferases like *MLL* or *EZH2*, or cohesines, that regulate the chromosomal interactions (Akalin *et al.*, 2012; Schoofs *et al.*, 2014). It was also reported that some oncogenic transcription factors (e.g. RUNX1-RUNX1T1 protein) may recruit DNMTs to their target sites and cause an abnormal methylation. On the contrary, other transcription factors are able to protect their binding sites from methylation (Gebhard *et al.*, 2010; Schoofs *et al.*, 2014). Thus, specific methylation losses were detected in patients with *NPM1*^{mut} and cooccurring mutations in *NPM1*, *DNMT3A*, and *FLT3*. Distinct methylation patterns were also found in AMLs with *CEBPA* or *RUNX1* mutations, and with gene fusions of transcription factors like CBF AML and APL (Figueroa *et al.*, 2010; Ley *et al.*, 2013; Gebhard *et al.*, 2019).

1.3.1.1 Prognostically significant changes in DNA methylation

Because of the great heterogeneity of AML, there is a constant search for novel biomarkers that would help to stratify the patients. DNA methylation is an attractive target mainly because it is a stable marker that is relatively easy to measure. Many studies use simple PCR-based methods and focus primarily on methylation levels in small regions, sometimes only one or a few loci, particularly in gene promoters (for example: Guo *et al.*, 2017; Hájková *et al.*, 2012; Liu *et al.*, 2017; Zhou *et al.*, 2017). With the advancement in high-throughput and NGS technologies, increasing number of researchers try to evaluate the methylation changes in a more complex way broadening the investigation to a large number of regions throughout the whole genome (Deneberg *et al.*, 2011; Luskin *et al.*, 2016; Kelly *et al.*, 2017; Li *et al.*, 2017).

Recently published studies evaluating the prognostic significance of DNA methylation aberrations in AML patients are summarized in **Table 4**. It is apparent that the outcome of a specific methylation change is highly dependent on its position in the genome. About half of the studies found out that higher methylation at a certain region is associated with better survival of patients (e.g. Deneberg *et al.*, 2011; Marcucci *et al.*, 2014; Qu *et al.*, 2017), however, other studies got to exactly opposite results when reporting that lower or no methylation at a certain locus is connected with favorable prognosis (for example

Treppendahl *et al.*, 2012; Jost *et al.*, 2014; Guo *et al.*, 2017). Not always there was a connection between DNA methylation and expression of the associated gene, so the biology behind some of these presumably prognostic significant changes is unclear. A few studies referred to the association with *HOX* genes because their expression is regulated epigenetically by polycomb and trithorax proteins and they play a central role in embryonic development and hematopoiesis (Deneberg *et al.*, 2011; Hájková *et al.*, 2012; Jost *et al.*, 2014). Some of the studies provided connection between genetic aberrations and DNA methylation changes (Marcucci *et al.*, 2014; Jost *et al.*, 2014; Kelly *et al.*, 2017; Liu *et al.*, 2017). Nevertheless, Li *et al.* (2016) highlighted in their work that genetic and epigenetic changes follow distinct patterns with different kinetics during the disease progression and thus should be evaluated separately as independent prognostic markers.

Unfortunately, none of these findings were implemented into clinical practice. The main reason is that there are no validation studies that would confirm or reject the clinical significance of reported DNA methylation changes with presumable prognostic impact. Another drawback is that these studies usually used different methods and so the results may also vary. For example, when comparing the broader studies that focused on promoter methylation, Figueroa *et al.* (2010) presented a methylation classifier based on 15 genes, Deneberg *et al.* (2011) discovered 62 CpGs associated with patients' survival, and Marcucci *et al.* (2014) identified 82 genes, methylation of which was associated with OS, and proposed a 7-gene panel for prognostic evaluation of AML patients. Surprisingly, there was no overlap between the identified prognostically significant DNA methylation changes among these three papers. On the other hand, when Qu *et al.* (2017) were searching for survival-associated methylation regions, they re-discovered some of the sites published by Marcucci *et al.* (2017).

Table 4 Summary of potential prognostically relevant changes in DNA methylation of AML patients published in recent studies

Author	Studied region/gene	Clinical significance	Notes
Figueroa *et al.* (2010)	Fifteen genes including transcription factors, protein-metabolism regulators, genes related to telomeres regulation, and signaling	Unique methylation classifier computed using supervised principal components (SuperPC) method predictive of OS ($p < 0.001$).	Aberrant methylation associated with downregulated expression.
Deneberg *et al.* (2011)	PcG (polycomb group proteins) target genes	Higher methylation associated with better disease-free survival ($p = 0.01$) and OS ($p = 0.001$) in CN-AML.	
Lin *et al.* (2011)	*CEBPA* distal promoter	Higher methylation was associated with longer OS ($p = 0.03$) in patients with normal karyotype and without *CEBPA*^{mut} and *NPM1*^{mut}.	
Hájková *et al.* (2012)	Promoters of tumor suppressor genes (*CDKN2B, ESR1, MYOD1, CALCA, SOCS1, CDH1*) and *HOX* genes	Hypermethylation of *SOCS1* promoter associated with better outcome ($p = 0.04$). Patients with smaller number of hypermethylated genes ($p = 0.012$) or with lower levels of cumulative DNA methylation value computed from methylation levels of all studied regions have worse OS ($p = 0.027$).	Studied negative impact of *HOX* genes and tumor suppressors' promoters hypomethylation caused by *DNMT3A*^{mut}.
Treppendahl *et al.* (2012)	*VTRNA2-1* promoter	Patients with hypermethylation had poorer survival ($p = 0.001$).	Methylation was inversely correlated with expression.
Marcucci *et al.* (2014)	DMRs in promoters of seven genes (*CD34, RHOC, SCRN1, F2RL1, FAM92A1, MIR155HG,* and *VWA8*)	High DMRs methylation associated with lower expression linked to higher CR rate, longer disease-free survival and OS ($p < 0.001$) in CN-AML.	*FLT3*-ITD and *DNMT3A* mutations associated with low methylation at DMRs, *NPM1* and *IDH* mutations associated with higher methylation at DMRs ($p < 0.001$).
Hájková *et al.* (2014)	*PBX3* (TAF1 binding site)	Lower methylation correlated with higher expression of *PBX3* that was associated with higher incidence of relapse ($p = 0.004$).	Newly discovered hypomethylation pattern specific to *CBFB-MYH11* fusion with corresponding gene overexpression.
Jost *et al.* (2014)	Promoter region of *DNMT3A*	Aberrant hypermethylation (> 10 %) detected in ~40 % of studied patients with shorter EFS and OS and poor or intermediate cytogenetic risk.	Higher methylation in the region were mostly observed in patients without *DNMT3A*^{mut} and was associated with moderate downregulation of *DNMT3A* transcription.
Božić *et al.* (2015)	1 CpG in *C1R* gene	Higher methylation (> 27 %) associated with longer OS ($p < 0.0001$).	Only moderate association of DNA methylation and expression of *C1R*.

Author	Studied region/gene	Clinical significance	Notes
Wertheim *et al.* (2015); Luskin *et al.* (2016)	Seventeen loci identified as having a prognostic significance for AML	Higher M-score (methylation statistic computed with the random forest method) associated with death (p = 0.01) and failure to achieve CR (p = 0.03).	
Li *et al.* (2016)	Promoter associated loci, each including four consecutive CpGs, with epigenetic allelic variance	Higher epigenetic allele burden (magnitude of epiallele shifting, gain or loss, across the genome) disrupts transcriptional regulation and is associated with shorter time to relapse (p = 0.008).	During disease progression, genetic and epigenetic allelic variations follow different kinetics and patterns.
Zhou *et al.* (2016)	*DLX4*	Patients with methylated *DLX4* presented lower CR rate (p = 0.001) and shorter OS (p = 0.003).	*DLX4* hypermethylation is negatively associated with expression.
Kelly *et al.* (2017)	Many CpG islands with aberrant methylation (CIMP), particularly *SCGB3A1*, *NPM2*, *CDKN2B*, and *OSCP1*	Patients with long survival (> 1-year, median OS = 90 months) had more aberrant methylation at studied CGIs (p = 0.02).	Patients with CIMP hypermethylation associated with IDH1/2^{mut} showed poorer OS, than patients with IDH1/2^{mut}-independent CIMP hypermethylation (p = 0.08).
Qu *et al.* (2017)	CGI shores of *LZTS2* and *NR6A1*	Hypomethylation in either of the two regions associated with worse OS (p < 0.001).	Studied on AML patients with normal karyotype.
Li *et al.* (2017)	*NKD2* promoter	Higher methylation correlated with lower expression of *NKD2* which was associated with shorter OS (p = 0.03) in CN-AML.	
Zhou *et al.* (2017)	*GPX3* promoter	MDS patients with *GPX3* methylation showed shorter OS (p = 0.01).	*GPX3* methylation increased during progression to AML.
Guo *et al.* (2017)	*SFRP1* and *SFRP2* promoter regions	Higher methylation associated with shorter OS (p = 0.03).	
Liu *et al.* (2017)	*RASSF1A* promoter	Hypermethylation connected with decreased relapse-free survival (p = 0.04).	Hypermethylation of *RASSF1A* associated with *ASXL1* mutations and decreased mRNA levels.
Šestáková *et al.* (2019)	*GZMB* enhancer	Hypermethylation associated with inferior OS (p = 0.03).	Concurrent presence of both *DNMT3A*^{mut} and *IDH1/2*^{mut} partially cancel out the opposite influence of these aberrations on DNA methylation resulting in a mixed methylation and hydroxymethylation profiles.

CGI – CpG island, CIMP – CGI methylator phenotype, CR – complete remission, DMR – differentially methylated region, EFS – event-free survival, MDS – myelodysplastic syndrome, OS – overall survival

1.3.2 DNA hydroxymethylation in AML

There are more DNA modifications resulting from DNA methylation. TET enzymes catalyze oxidation of the methyl group subsequently resulting in 5′-hydroxymethylcytosine, 5′-formycytosine, and 5′-carboxylcytosine. All of these forms of oxidized 5′-methylcytosine are stable epigenetic marks that inhibit DNMT1 causing a passive demethylation of the locus. 5′-formycytosine, and 5′-carboxylcytosine also trigger the active DNA demethylation pathway mediated by base excision repair (Ko *et al.*, 2015).

The biological function of these 5-mC derivatives is yet to be established but a role in the gene expression modulation has already been proposed. It was shown that these DNA modifications can regulate the recruitment and binding of polycomb repressive complexes and can reverse the transcriptional silencing (Fong *et al.*, 2014; Wouters & Delwel, 2016). Furthermore, it was suggested, that DNA hydroxymethylation can play an important role in shaping the epigenetic landscape of cancer genomes and that TET enzymes can act as both tumor suppressors and promoters of a malignancy (Jeschke *et al.*, 2016). A study evaluating DNA hydroxymethylation in AML patients found that 5-hmC levels were related to increased expression of associated genes. Moreover, this correlation was even more significant than between DNA methylation and downregulation of the gene (Rampal *et al.*, 2014). Another research found an association between overall 5-hmC levels and AML prognosis, showing that higher levels of 5-hmC correlate with inferior survival of patients (Kroeze *et al.*, 2014).

1.3.3 Other epigenetic changes in AML

In previous chapters, a role of histone modifications in AML epigenetic landscape was already mentioned. Histones (H1, H2A, H2B, H3, and H4) are basic globular proteins that form nucleosomes, key components in DNA organization. Loose *N*-terminal trails of histones can be subjected to a variety of modifications like methylation, acetylation, phosphorylation, ubiquitylation, etc. These alterations regulate the associated DNA region. Open chromatin structure allows gene expression and is marked by histone acetylation and methylation of certain histone lysine residues (H3K4me3, H3K36me3 or H3K79me2). Methylation of histone arginine residues also promotes transcriptional activation. In contrast, trimethylation of other H3 lysine residues (H3K9me3, H3K27me3) is a marker of gene repression which is also promoted by histone deacetylases (Hájková, 2015).

The DNA and histone methylation mechanisms are tightly linked by various enzyme interactions. For example, it was found that EZH2 mediated methylation of H3K27, a repression mark formed by polycomb complex, may recruit DNA methyltransferases

in cancer cells (Schlesinger *et al.*, 2006). Also, the oncometabolite produced by mutated IDH1/2 enzymes, 2-hydroxyglutarate, inhibits a group of lysine demethylases containing Jumonji domain, leading to altered gene expression and impaired cell differentiation (Fong *et al.*, 2014). Furthermore, lysine methyltransferases (mainly KMT2A and EZH2) were found to be recurrently mutated in AML and these patients exhibit specific DNA methylation changes (Schoofs *et al.*, 2014; Wouters & Delwel, 2016).

The histone modifications themselves have an impact on AML pathogenesis. Chromatin-binding protein ASXL1 (described in chapter 1.2.2.1) should be mentioned because its interaction with EZH2, which is particularly important in AML development, leads to a global decrease of methylation at H3K27 residues and impaired upregulation of genes, including *HOX* genes (Fong *et al.*, 2014). A recent study also showed that arginine methyltransferase PRM1 promotes the expansion of *FLT3*-ITD leukemic cells and thus PRMT inhibitors were proposed to enhance the AML therapy (He *et al.*, 2019). Histone deacetylases were not found to be mutated in AML. Nevertheless, it was shown that these enzymes can be recruited by some oncoproteins, for instance the PML-RARA fusion protein, resulting in aberrant gene silencing (Abdel-Wahab & Levine, 2013; Fong *et al.*, 2014).

RNA-based mechanisms of gene silencing are also commonly classified as epigenetic regulations. RNA interference is a mechanism where short non-coding RNAs (miRNA, siRNA) cause degradation of target mRNA. Investigations of miRNAs in AML found for instance that certain miRNAs expression is associated with cytogenetics and *FLT3* mutations and that higher expression of *miR-191* and *miR-199a* are linked to a worse outcome in AML patients (Garzon *et al.*, 2008). Later studies discovered a tumor suppressing *miR-29b* (Garzon *et al.*, 2009), and *miR-29a* and *miR-142-3p* decreased expression of which is involved in the development of AML (Wang *et al.*, 2012). Long non-coding RNAs (lncRNA), involved in many biological processes such as imprinting, epigenetic regulation and apoptosis, were also found to have a role in AML pathogenesis. Various genetic lesions of AML were characterized by distinct profiles of lncRNAs and a subset of lncRNAs that strongly correlated with patients' survival was also discovered (Garzon *et al.*, 2014). It was further found that upregulated lncRNAs in AML are associated with lower DNA methylation levels, and that lncRNA *LOC285758* enhances the expression of histone deacetylase 2 and thus regulates the proliferation of AML cells and is linked to poor prognosis (Lei *et al.*, 2018).

1.4 Methods for genetic and DNA methylation studies

1.4.1 Next Generation sequencing in AML

Next generation sequencing (NGS) is a high-throughput method of massive parallel sequencing that allows for rapid and precise evaluation of multiple samples. Thanks to these advances in sequencing technologies in the last decade, an immense amount of new knowledge about the mutational landscape of AML was acquired that influenced AML classification, prognostic stratification, and even treatment choices and response assessment (Ley *et al.*, 2013; Papaemmanuil *et al.*, 2016; Folta *et al.*, 2019).

There are various types of NGS technology. Thermo Fisher Scientific (USA) Ion proton system is based on semiconductors that measure the slight pH change that occurs when hydrogen ion is released during the attachment of a new base to the newly synthesized DNA strand. PacBio System developed by Pacific Bioscience (USA) is able to measure the fluorescent signal of each base attached to a single DNA molecule in real-time. Real-time sequencing of a single molecule is also used in Oxford Nanopore Technology (UK) that measures the electrical current of each base on a DNA strand passing through a nanopore placed on a synthetic membrane. However, the most commonly used technology is sequencing by synthesis utilizing fluorescently labeled reversible terminators developed by the Illumina company (USA) (Leisch *et al.*, 2019).

The price of NGS depends mainly on the amount of genome that is sequenced. Whole genome sequencing (WGS) or whole exome sequencing (WES) are much more expensive than targeted sequencing that focuses only on certain chosen regions of DNA. Targeted sequencing is currently the most popular method and begins to be utilized in common clinical practice. There is still no consensus about the number and composition of genes that should be included in a myeloid sequencing panel. Nevertheless, there are several commercially available panels targeting genes most relevant for AML pathogenesis, e.g. AmpliSeq® Myeloid Sequencing Panel (Illumina), see **Table 5**, SureSeq myPanel™ NGS Custom AML (Oxford Gene Technology), and Human Myeloid Neoplasms Panel (Qiagen). It is also foreseen that once the costs and turnaround times of NGS technology are further reduced, whole-exome or even whole-genome sequencing will become the standard for AML characterization at diagnosis (Leisch *et al.*, 2019; Levine & Valk, 2019).

Table 5 AmpliSeq™ for Illumina Myeloid Panel gene list

Hotspot genes (n=23)									
ABL1	BRAF	CBL	CSF3R	DNMT3A	FLT3	GATA2	HRAS	IDH1	IDH2
JAK2	KIT	KRAS	MPL	MYD88	NPM1	NRAS	PTPN11	SETBP1	SF3B1
SRSF2	U2AF1	WT1							
Full genes (n = 17)									
ASXL1	BCOR	CALR	CEBPA	ETV6	EZH2	IKZF1	NF1	PHF6	PRPF8
RB1	RUNX1	SH2B3	STAG2	TET2	TP53	ZRSR2			
Fusion driver genes (n = 29)									
ABL1	ALK	BCL2	BRAF	CCND1	CREBBP	EGFR	ETV6	FGFR1	FGFR2
FUS	HMGA2	JAK2	KMT2A (MLL)	MECOM	MET	MLLT10	MLLT3	MYBL1	MYH11
NTRK3	NUP214	PDGFRA	PDGFRB	RARA	RBM15	RUNX1	TCF3	TFE3	
Expression genes (n = 5)			Expression control genes (n = 5)						
BAALC	MECOM	MYC	SMC1A	WT1	EIF2B1	FBXW2	PSMB2	PUM1	TRIM27

1.4.2 Methods for DNA methylation analysis

Methods for epigenetic studies are also rapidly evolving. A few approaches have been developed to enrich the analyzed DNA regions of CpG sites that could possess a differential methylation: immunoprecipitation method based on antibody against 5-mC, a pull down of methylated regions using 5-mC binding protein (MBD), or the use of restriction enzymes that can be both methylation sensitive or insensitive and cleave a CpG site (for example HpaII, a methylation-sensitive enzyme that cuts CCGG sequence, and MspI, that cuts the same site but regardless of its methylation status) (Kurdyukov & Bullock, 2016; Šestáková *et al.*, 2019). With these preparatory steps, a pool of CpG rich regions, mostly from CGIs, is obtained. Nevertheless, the majority of studies utilize bisulfite (BS) conversion of DNA as an initial step of DNA methylation assessment. Sodium bisulfite is used to mediate the deamination of cytosines into uraciles while methylated cytosines remain intact (Hayatsu *et al.*, 1970). Usually, PCR amplification follows the conversion and thus converted residues are changed to thymines while methylcytosines remain cytosines. This method originally required great amount of DNA and was very damaging for the sample. Nowadays, there are commercial kits that able to convert as little as 100 pg of DNA in less than 2 hours and guarantee 99 % conversion efficiency (Šestáková *et al.*, 2019). The main benefit of this approach is that any region of DNA or even a whole genome can be converted and examined.

1.4.2.1 High-throughput methods

Before NGS became generally available, other methods allowing DNA methylation analysis of multiple regions were used. HELP (HpaII tiny fragment enrichment by ligation-mediated

PCR) assay consists of MspI and HpaII digestion of genomic DNA and subsequent ligation and PCR amplification of resulting fragments (Khulan, 2006). Fluorescently labeled amplicons are then hybridized to custom-made oligonucleotide microarrays and signals from MspI and HpaII digestions are compared. This method was further optimized into MELP (Microsphere HpaII tiny fragment enrichment by ligation-mediated PCR) where the labeled amplicons are hybridized to microspheres with distinct fluorescent properties and then subjected to flow cytometry. The microsphere fluorescence identifies the locus and signal intensity of the PCR products reflects the methylation level (Wertheim *et al.*, 2014). The authors further reduced the time for this assay by altering the linker oligonucleotides and called the improved assay xMELP (Expedited microsphere HpaII small fragment enrichment by ligation-mediated PCR) (Wertheim *et al.*, 2015).

A variety of microarrays was also developed for DNA methylation studies. Originally, the arrays were designed for hybridization of samples after the immunoprecipitation step, e.g. Human CpG Island Microarray Kit (Agilent). However, these kinds of chips were replaced with microarrays investigating the methylation status of BS converted samples, developed mostly by Illumina company. For example, Infinium HumanMethylation450 Bead Chip was used in The Cancer Genome Atlas 2013 study of large cohort of AML patients (Ley *et al.*, 2013). Illumina further extended this microarray and presented Infinium MethylationEPIC Bead Chip that covers over 850,000 CpGs targeting CGIs, UTR regions and even CpGs inside gene bodies. The chip uses two types of probes, type I uses two probes for one locus, one for methylated and one for unmethylated allele. Type II probes use single probe for both methylated and unmethylated allele which are distinguished by a color mark that is attached to analyzed DNA fragments during sample preparation (Pidsley *et al.*, 2016). Still, the use of microarrays is quite expensive because of the limited capacity of the chip (12 samples) (Kurdyukov & Bullock, 2016). Therefore, these techniques were shaded by rapidly evolving NGS approaches.

Early NGS studies of epigenetic landscape in AML preferentially used the enrichment methods (restriction enzymes, immunoprecipitation) as the first step to focus on aberrant methylation at CGIs. Reduced representation bisulfite sequencing (RRBS) uses MspI restriction fragments that are linked to adapters, treated with bisulfite, amplified, and sequenced (Meissner, 2005). This method was further adjusted to enhanced RRBS (ERRBS) by selecting longer MspI fragments to capture more regions, mainly beyond CGIs (Akalin *et al.*, 2012). Nowadays, with the focus aimed preferentially at CpGs outside CGIs, whole genome bisulfite sequencing (WGBS) is the method of choice for complete

examination of the genome. There, the DNA is only bisulfite converted, amplified with random primers and sequenced. However, BS sequencing of the whole genome is not only expensive, but also extremely challenging for bioinformatic analysis, mainly the mapping of acquired data, because bisulfite conversion markedly decreases the complexity of reads. Targeted bisulfite sequencing is thus more common for instance with SureSelect Human Methyl-Seq (Agilent) or SeqCap Epi Enrichment Kit (Roche) technologies. It is also possible to design and purchase custom made probes for the capture of specific sequences. In the future, single molecule sequencing approaches may further simplify the methylation assessment. For example, the nanopore technology (mentioned in chapter 1.4.1) could readily distinguish cytosine and methylated cytosine in a sequence without any previous DNA treatment (Kurdyukov & Bullock, 2016; Simpson *et al.*, 2017).

1.4.2.2 DNA methylation analysis of specific regions

Beside the high-throughput technologies, there is a number of methods for measuring the DNA methylation level at certain loci or short region. These techniques are indispensable for validation of the data from NGS or microarrays and are usually quick, cost effective, sensitive, and suitable for measurement of large number of samples (Šestáková *et al.*, 2019).

One of the fastest methods employs the methylation specific and unspecific restriction endonucleases (MSRE). After cleavage, the number of resulting fragments is measured by quantitative PCR. This approach enables the assessment of methylation level only at CpGs inside specific restriction sites which is the main drawback (Itoi *et al.*, 2007). There are commercially available kits for this method, e.g. OneStep qMethyl kit (Zymo Research) (Kurdyukov & Bullock, 2016).

Other methods require bisulfite conversion. The golden standard of DNA methylation measurements is bisulfite sequencing, where selected bisulfite converted DNA region is cloned into a vector and transformed into bacteria. At least ten bacterial colonies are then harvested and their vectors are sequenced for the DNA methylation assessment (Frommer *et al.*, 1992). Another approach with base-resolution is pyrosequencing, where a specific region of BS converted DNA is amplified via PCR and then sequenced. Pyrosequencing uses a mix of enzymes that produce light in every sequencing cycle when the specific base is attached to immobilized template strand. The methylation percentage is counted from the signal ratio of attached cytosines and thymines at certain CpG loci (Tost & Gut, 2007). The rest of the methods lack the base resolution and are able to measure only the average methylation of the whole region of interest. Methylation specific

high-resolution melting analysis (MS-HRM) is based on different melting temperatures (T_m) of methylated and unmethylated DNA. The studied region is amplified with PCR and then a melting analysis is performed with small ramping (0.1°C). Melting curves can be distinguished because of the different T_m between CG base pair of methylated cytosine and AT base pair resulting from converted unmethylated cytosine (Wojdacz & Dobrovic, 2007). Very common method is methylation-specific PCR (MS-PCR) which requires two sets of primers that span the region of interest, one for methylated allele and the other for unmethylated allele (Herman *et al.*, 1996). More quantitative modification of MS-PCR is called MethylLight and uses methylation specific probes in quantitative PCR. The probes are designed for any combination of the methylated state inside the selected region (methylated/unmethylated allele of each CpG) which increases the accuracy of the DNA methylation assessment (Eads *et al.*, 2000).

1.4.3 Methods for DNA hydroxymethylation detection

As described in chapter 1.3.2, DNA hydroxymethylation may too play a role in AML pathogenesis and thus methods for the detection and measurement of 5-hmC should be mentioned. Global DNA hydroxymethylation levels can be measured by liquid chromatography of DNA digested to single nucleotides and linked to electrospray ionization tandem mass spectrometry (LC-MS/MS). LC-MS/MS can distinguish all bases and their modifications and measure their amount in the sample (Le *et al.*, 2011). A few methods were developed for the investigation of 5-hmC in specific regions. Digestion-based techniques use a glycosyltransferase in the first step. Only hydroxymethyl cytosines are glycosylated and thus resistant to subsequent MspI cleavage. The resulting DNA fragments can be analyzed via PCR or NGS. 5-hmC antibodies are also available for the enrichment of hydroxymethylated sites prior to sequencing (Kurdyukov & Bullock, 2016). Furthermore, an adjusted BS-based method called oxidative bisulfite conversion (oxBS) was developed, where an additional oxidation step is performed prior to the BS conversion itself. The hydroxymethylated cytosines are oxidized to formylcytosines which are subsequently converted by BS to thymines together with the unmethylated cytosines. The location of 5-hmC is extracted from a comparison of BS and oxBS converted samples analyzed either via microarray or NGS (Booth *et al.*, 2012).

2 AIMS OF THE THESIS

With the advancements in high-throughput sequencing, there was a great progress in mapping of the genetic landscape of AML that contributed to a more accurate risk classification (Döhner *et al.*, 2017). However, AML is a highly heterogenous malignancy and patients with this disease would still benefit from a better prognostic stratification. It is well known that nearly half of AML patients harbor mutations in genes that regulate epigenetic mechanisms, especially DNA methylation (Ley *et al.*, 2013). It is therefore understandable to assume that DNA methylation by itself may affect the prognosis of AML patients. Indeed, many studies addressed and confirmed the prognostic potential of various DNA methylation changes. However, none of the reported prognostically significant DNA methylation aberrations were implemented into clinical practice. The reason is a prevailing inconsistency of these studies, mainly in used methods and studied regions, and also a lack of robust validation studies.

The aims of this thesis were to deeper investigate the DNA methylation changes in AML patients with defined genetic background, to assess the suitability of particular DNA methylation validation techniques, and to evaluate potentially predictive DNA methylation changes using a comprehensive approach.

Firstly, we focused on the overall methylation profiles of patients with mutations in either *DNMT3A* or *IDH1/2* genes and their combinations. These mutations have an opposing effect on DNA methylation and we were therefore curious how the methylation profiles would differ. Secondly, we aimed for the evaluation of DNA methylation validation techniques since choosing the appropriate methodology is a critical aspect of validation studies. Finally, we wanted to develop a new complex NGS-based approach for the overall evaluation of DNA methylation changes that were previously described as having a prognostic significance. We hoped that the results from a custom DNA methylation sequencing panel would both validate the results previously published by other authors and help to better stratify the AML patients, mostly those within intermediate risk group whose outcome and therapeutic strategy is still not unequivocal.

3 MATERIAL AND METHODS

3.1 DNA methylation and hydroxymethylation changes in AML patients with mutations in *DNMT3A*, *IDH1/2* or their combinations

3.1.1 Patients

De novo CN-AML patients (n = 24) were chosen from a collaborative mutational study of 58 non-APL intensively treated AML patients examined for their mutational status by ClearSeq AML kit (Agilent Technologies, Santa Clara, CA, USA) (Folta *et al.*, 2019). Selected samples were divided according to their mutations into 5 groups: *IDH1*^{mut} (n = 4), *IDH2*^{mut} (n = 4), *DNMT3A*^{mut} (n = 8), *DNMT3A&IDH1*^{mut} (n = 4), and *DNMT3A&IDH2*^{mut} (n = 4). Summary of patients' mutational background is shown in **Table 6**. We preferentially picked samples with the same type of mutation in the three genes and with higher VAF. In these patients, overall hydroxy-/methylation and expression profiles were measured with arrays. Gene expression and DNA methylation of specific genes was further measured in a large cohort of diagnostic AML samples (n = 104, patients' clinical characteristics are provided in **supplementary Table 1** in section 9).

This study was approved by the Institutional Ethics Committee and all patients gave their full consent. The research also conforms with The Code of Ethics of the World Medical Association.

Table 6 Characteristics of CN-AML patients investigated with methylation and expression arrays

Sample	Sex	Blasts [%]	NPM1 activated signaling	activated signaling				DNA methylation				transcription factor	spliceosome
			NPM1 (%VAF)	**FLT3-ITD length**	**FLT3-TKD (%VAF)**	**NRAS (%VAF)**	**CBL (%VAF)**	**DNMT3A (%VAF)**	**IDH1 (%VAF)**	**IDH2 (%VAF)**	**TET2 (%VAF)**	**CEBPA (%VAF)**	**SRSF2 (%VAF)**
DNMT3Amut_1	F	58	A 37.2%	+30bp	negative	negative	negative	**R882H 46.4%**	negative	negative	negative	negative	negative
DNMT3Amut_2	F	53	negative	+51bp	negative	negative	negative	**R882H 43.7%**	negative	negative	negative	negative	negative
DNMT3Amut_3	F	58	A 38.8%	+21bp	negative	negative	negative	**R882H 46.8%**	negative	negative	negative	negative	negative
DNMT3Amut_4	F	71	A 40.5%	+57bp	negative	negative	negative	**R882C 46.6%**	negative	negative	negative	negative	negative
DNMT3Amut_5	F	52	negative	+156bp	negative	negative	negative	**R882H 46.4%**	negative	negative	M1701I 46.8%	negative	negative
DNMT3Amut_6	M	54	A 39.0%	+81bp	negative	negative	negative	**R882H 43.0%**	negative	negative	negative	negative	negative
DNMT3Amut_7	M	65	negative	negative	negative	negative	negative	**R882H 44.7%**	negative	negative	negative	negative	negative
DNMT3Amut_8	F	64	negative	+42bp	negative	negative	negative	**R882H 49.8%**	negative	negative	K536fs 49.6%	negative	negative
IDH1mut_1	F	84	A 39.0%	negative	negative	G12V 21.2%	Y368D 2.3%	negative	**R132C 42.0%**	negative	negative	negative	negative
IDH1mut_2	M	89	D 43.4%	negative	D835V 29.5% D835Y 7.2%	G12D 2.3%	negative	negative	**R132H 39.4%**	negative	negative	negative	negative
IDH1mut_3	F	65	A 10.8%	+57bp	negative	Q61K 36.8%	negative	negative	**R132H 37.8%**	negative	negative	negative	negative
IDH1mut_4	M	65	negative	negative	negative	negative	negative	negative	**R132C 33.4%**	negative	negative	negative	negative
IDH2mut_1	F	90	A 42.3%	+18bp	negative	negative	negative	negative	negative	**R140Q 44.5%**	negative	negative	negative
IDH2mut_2	M	64	A 43.5%	negative	negative	negative	negative	negative	negative	**R140Q 42.8%**	negative	negative	P95delinsRP 35.3%
IDH2mut_3	M	84	A 41.5%	negative	negative	negative	negative	negative	negative	**R140Q 46.1%**	negative	negative	negative
IDH2mut_4	M	75	A 37.3%	negative	negative	negative	negative	negative	negative	**R140Q 48.1%**	negative	negative	negative
DNMT3A&IDH1mut_1	M	50	A 42.0%	negative	D835Y 20.2%	negative	negative	**R882H 45.2%**	**R82K 49.5%**	negative	negative	P197delinsHPP 21.0%	negative
DNMT3A&IDH1mut_2	F	78	A 44.5%	negative	negative	G13D 2.0% Q61K 2.9%	negative	**R882C 45.3%**	**R132S 47.5%**	negative	negative	negative	negative
DNMT3A&IDH1mut_3	F	86	negative	negative	negative	negative	negative	**R882H 32.4%**	**R132C 29.9%**	negative	negative	negative	negative
DNMT3A&IDH1mut_4	F	62	A 37.3%	negative	negative	G13D 38.0%	negative	**R882H 44.7%**	**R132H 35.0%**	negative	negative	negative	negative
DNMT3A&IDH2mut_1	M	86	negative	+36bp	negative	negative	negative	**R882H 42.7%**	negative	**R140Q 46.4%**	negative	negative	negative
DNMT3A&IDH2mut_2	M	50	B 18.9%	negative	negative	negative	C416S 4.9% R420Q 67.6%	**R882C 40.5%**	negative	**R140Q 36.6%**	negative	negative	negative
DNMT3A&IDH2mut_3	M	57	negative	negative	negative	negative	negative	**R882C 45.3%**	negative	**R140Q 43.4%**	negative	negative	negative
DNMT3A&IDH2mut_4	M	64	negative	negative	negative	negative	negative	**R882P 29.0%**	negative	**R140Q 34.0%**	negative	negative	negative

3.1.2 Samples' preparation

In case of AML samples investigated with arrays, RNA and DNA were extracted with AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) from peripheral blood leukocytes lysed in RLT buffer.

In the larger AML cohort where methylation and expression of specific genes were assessed, DNA was extracted either from peripheral blood with MagCore system (RBCBioscience, New Taipei City, Tchaj-wan) or with AllPrep DNA/RNA Mini Kit from RLT lysates. RNA was isolated with TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA) from peripheral blood leukocytes. Reverse transcription was performed with M-MLV Reverse Transcriptase (Promega, Madison, WI, USA).

Ficoll gradient centrifugation (Histopaque, Sigma-Aldrich, St. Louis, MO, USA) was used for isolation of mononuclear cells (MNC) of selected AML patients (from diagnostic whole blood samples) or healthy blood donors (from buffy coats). CD34+/CD117+ cells were isolated from MNC using magnetic separation with MicroBeads kit (Miltenyi Biotec, Bergish Gladbach, Germany). DNA was subsequently extracted with MagCore instrument (RBCBioscience).

3.1.3 Methylation and hydroxymethylation assessment with arrays

TrueMethyl-Seq kit (CEGX, Cambridge, UK) was used for both BS and oxBS conversion of 1.2 μg of DNA from each sample (AML: n = 24, listed in **Table 6**; healthy donors: n = 4). BS and oxBS-converted DNA (140-250 ng) was subsequently investigated using Infinium MethylationEPIC BeadChip (Illumina, San Diego, CA, USA). Raw IDAT files were analyzed with R package ChAMP and β-values that correspond to methylation percentage (range 0-1), were acquired for each investigated locus. The batch effect between individual arrays was corrected using ComBat function (Johnson *et al.*, 2007) and β-values were normalized via Functional normalization (Fortin *et al.*, 2014).

OxBS β-values were used for DNA methylation assessment. DNA hydroxymethylation levels were computed from the difference between BS and oxBS β-values ($\Delta\beta_{BS\text{-}oxBS}$). A threshold was set based on the 95th percentile of negative $\Delta\beta_{BS\text{-}oxBS}$ values across all examined samples, as recommended in previous publication (Lunnon *et al.*, 2016) and lower $\Delta\beta_{BS\text{-}oxBS}$ values were considered as not hydroxymethylated. R package pvclust (Suzuki & Shimodaira, 2006) was used for all hierarchical clustering analyses.

3.1.4 Bisulfite conversion

EZ DNA Methylation-Lightning Kit (Zymo Research, Irvine, CA, USA) was used for DNA (~ 500 ng) treatment. Concentration of converted DNA was measured with NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific).

3.1.5 Pyrosequencing

PyroMark PCR kit (Qiagen) was used for the initial PCR amplification of oxBS and BS converted DNA (10-20 ng). The PCR reaction and conditions were kept according to the manufacturer's protocol with 25 µl final volume and 0.2 µM final concentration of forward and reverse biotinylated primers. When the universal biotinylated primer was used, the final concentration of primers was 0.2 µM of forward and universal biotinylated primer and 0.04 µM of reverse tailed primer. Primers' sequences and used annealing temperatures (T_{ann}) are shown in **Table 7**. Quality of acquired amplicons (1 µl of PCR reaction) was checked using 2% agarose gel electrophoresis. Pyrosequencing was performed on PyroMark Q24 instrument (Qiagen). Two steps of the original manufacturer's protocol (User Manual 01/2016) were optimized: 2 µl of sepharose-coated streptavidin beads were added in step 5.3.3.2 and the agitation was prolonged to 20 min (step 5.3.3.6).

Table 7 Primers used for DNA hydroxymethylation and methylation validation by pyrosequencing

Gene	Primer	Sequence	T_{ann} [°C]
GZMB	Forward	5'- TATAAATAGAGTTGTTTTGGTG -3'	56
	Reverse biotinylated	5' biotin-AAACCATCATCTTCTCTAATAT -3'	
	Sequencing	5'- ATTGAGGTTTGGATGTTTTA -3'	
MYB	Forward	5'- GAGGTAGTTTATTAGATTTTG -3'	50
	Reverse biotinylated	5' biotin- TAATATATACCATCATCACC -3'	
	Sequencing	5'- TGTTTATTTTGAAGTTGTTG -3'	
CHFR	Universal biotinylated	5' biotin-ATCT*GTGCCGAGGCTCAGGC* -3'	
	Forward	5'- TAAGAATYGGTGGGTAGAATAT -3'	54
	Reverse - tailed	5'- *GTGCCGAGGCTCAGGC*ATCTCCRCAAATATAAATCC -3'	
	Sequencing	5'- TTTATTTTAYGGAAAAATTTGGAG -3'	
RNF216	Forward	5'- AGTTAATTTAGTTGAAATGTTAGGTTT -3'	54
	Reverse - tailed	5'- *GTGCCGAGGCTCAGGC*CTCTTTTCTTCTCACAAATTAAAA -3'	
	Sequencing	5'- AGTTAATTTAGTTGAAATGTTAGGTTT -3'	

3.1.6 Gene expression analyses

HumanHT-12 v4 Expression BeadChip (Illumina) was used to investigate the overall gene expression of samples (AML: n = 24, listed in **Table 6**; healthy donors: n = 6). Data analysis was performed in R with packages limma, sva and Biobase. Background correction

by NormExp and quantile normalization were applied on raw IDAT files. ComBat function was used for batch effect correction (Johnson *et al.*, 2007).

TaqMan Gene Expression Assays (Life Technologies, Carlsbad, MA, USA) were used to measure the expression levels of *GZMB* (Hs188051_m1), *CHFR* (Hs00943495_m1), and reference gene *GAPDH* (Hs02758991_g1) in the large cohort of AML samples (n = 104) and healthy donors (n = 10). The amplification was performed with TaqMan Universal MasterMix II (Life Technologies) with 10 µl final volume. The expression was measured on StepOnePlus Real-Time PCR System (Life Technologies) with conditions according to the manufacturer's protocol.

3.1.7 Statistics

The overall and event-free survivals were estimated by Kaplan-Meier curves with two-sided logrank test in GraphPad Prism 7 (GraphPad Software, La Jolla, CA, USA). SPSS software (IBM, Armonk, NY, USA) was used for Cox regression analysis. The correlation of DNA hydroxy-/methylation and expression data, and hydroxy-/methylation validation were performed in Microsoft Excel, F-test of overall significance and Pearson correlation coefficient were used.

3.1.8 Gene ontology analysis

Bed files with positions of differentially methylated sites were submitted to online tools GREAT (http://great.stanford.edu/public/html/) (McLean *et al.*, 2010) and Enrich (https://amp.pharm.mssm.edu/Enrichr/) (Kuleshov *et al.*, 2016).

3.1.9 DNA hydroxymethylation assessment via mass spectrometry

DNA (1 µg) of AML samples at diagnosis (n = 40) and two healthy donors' samples were first cleaved by DNA Degradase Plus (Zymo Research) according to manufacturer's protocol. The reaction was stopped with 0.1% formic acid added to 150 µl final volume of the mix. The mix was then filtered with centrifugal filters Microcon-10kDa with Ultracel-10 membrane (Sigma-Aldrich) and evaporated to ~ 3 µl volume in vacuum concentrator Speedvac SPD111V (Thermo Fisher Scientific). Samples were analyzed using LC-MS/MS with Shimadzu Prominence chromatograph (Shimadzu, Kyoto, Japan) and mass spectrometer QTRAP 4000 (Sciex, Framingham, MA, USA) using selected reaction monitoring mode. External calibration using synthetic standards was performed prior to the analysis.

3.2 DNA methylation validation methods

3.2.1 Samples and DNA standards

Mononuclear cells of healthy donors (n = 10) were harvested from buffy coats using Ficoll gradient centrifugation (Histopaque, Sigma-Aldrich). DNA was isolated with MagCore instrument (RBCBioscience). The blood donors provided their full consent and the study was approved by the Institutional Ethics Committee.

Human Non-methylated & Methylated DNA Set (Zymo Research) was utilized as unmethylated and methylated DNA standards.

3.2.2 Characterization of analyzed CpGs

Three CpGs with different levels of DNA methylation were chosen based on data acquired in a previous project (chapter 4.1) from healthy donors' samples analyzed on Infinium MethylationEPIC BeadChip (Illumina). Basic characteristic of selected loci is summarized in **Table 8**. Furthermore, these CpGs are placed within CCGG sequence and thus can be cleaved by MSRE.

Table 8 Characteristics of selected CpG sites with distinct methylation levels

Locus name	BeadChip probe ID	Cytosine location (hg 19)		Average beta value for samples measured with BeadChip
		Chromosome	Position	
M	cg24337108	10	11797422	> 0.99
IM	cg25722983	1	36840028	from 0.45 to 0.55
U	cg09655782	4	57333859	< 0.1

M – methylated, IM – intermediately methylated, U – unmethylated, BeadChip -Infinium MethylationEPIC BeadChip (Illumina), Beta value – a number corresponding to methylation percentage

3.2.3 MSRE analysis

For MSRE analysis, OneStep qMethyl Kit (Zymo Research) was utilized. Following the manufacturer's protocol, DNA (20 ng) of each sample was processed through Reference and Test reactions. Annealing temperature was set to 60 °C in the PCR step, and the time of annealing was shortened to 45 s. We used Rotor-Gene Q 2plex HRM Platform (Qiagen) for the measurements.

3.2.4 Bisulfite conversion

EZ DNA Methylation-Lightning Kit (Zymo Research) was used for the BS conversion. About 500 ng of DNA was treated and the concentration of converted DNA was measured with

NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific). For the MS-HRM and quantitative MS-PCR (qMSP) experiments, the concentration of treated DNA was adjusted to 10 $ng \cdot \mu l^{-1}$.

3.2.5 Primer design

Online software Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) was used to design primers for MSRE analysis. Methyl Primer Express Software v1.0 (Thermo Fisher Scientific) was employed when designing primers for methods examining BS converted DNA. Sequences and characteristics of all used primers are summarized in **Table 9**. **Figure 3** shows the three studied regions together with the positions of used primer pairs.

Table 9 Sequences and characteristics of used primers

Primer type	Forward/ sequencing primer	Reverse primer	T_{ann} [°C]	Product length [bp]
M pyrosequencing	GGTAGGAGGATGGTTTGAATT/ GGTGGAAATGAAGTAGGTGTGTTTG	*GTGCCGAGGCTCAGGC*AACACTACTCTTACCAAAACAACC	60	373/227
IM pyrosequencing	GTTAAGGGGGTGTATTTTAGAGA/ GGTAGAGAGAAGTTTTTTTTGTAGG	*GTGCCGAGGCTCAGGC*CTTAACTACTTTCCCAAACTACCT	58	399/339
U pyrosequencing	GGGGGGGTGTTAGTATTTG/ TTAGTATTTGYGTTGTGGAGTG	*GTGCCGAGGCTCAGGC*CCAAACTAACCTAATAAAACC	58	300/290
Universal biotinylated primer	5'biotin-*TCTGTGCCGAGGCTCAGGC*			
M MSRE	TTTTCTGTGACCTCCTTTGG	CAGTGTGACTGCTGGTGAAG	60	243
N MSRE	GCAATAGGCGTTAATGTCGT	AGGAGTGGCAAAAGAGGACT	60	199
U MSRE	CGCTTAGCAATCATCGACTT	GAAACAGGCCGCATCCTC	60	265
M MSP Met	GTATATTCGGAATTATTTCGTTTTC	AATTAACAACCGACAACCG	56	72
M MSP Unm	GATGTATATTTGGAATTATTTTGTTTTT	AATTAACAACCAACAACCA	56	75
IM MSP Met	CGGTTTTTATAGTTTTGAATTAGATC	TTATTTATTATCACATCAACTACTTCCG	58	166
IM MSP Unm	ATTGGTTTTTATAGTTTTGAATTAGATT	TTATTTATTATCACATCAACTACTTCCA	58	168
U MSP Met	CGTTGTGGAGTGAAGTGAATC	ACCGAACGAACAATAAACGAA	54	210
U MSP Unm	TGTGTTGTGGAGTGAAGTGAATT	ACCAAACAAACAATAAACAAAAAA	54	212
M HRM	TTGGGTGGAAATGAAGTAGGTGTG	CCAAACCATTAACCATAACAATA	54-58[1]	94
IM HRM	TTTGGGGAAAAAATATATGGAGTT	CTACTAATAAAACCCTTTACTCCCA	54-58[1]	90
U HRM	TTAGTATTTGYGTTGTGGAGTG	CCRACACTTACTCTTATTAACRATC	54-58[1]	93
M HRM Wojdacz	**CG**GGGGGGTGTTAGTATTTG	CC**CG**ACACTTACTCTTATTAACRATC	55	110
U HRM Wojdacz	T**CG**TGTTTTTTTTTGGGTGGAAATG	**GCG**ACCAAACCATTAACCATAACA	55	104

[1]For HRM experiments T_{ann} was 55°C, in MSP experiments T_{ann} of MSP primers was used
M – methylated locus, IM – intermediately methylated locus, U – unmethylated locus, MSP – methylation specific PCR, Met – primers for methylated DNA sequence, Unm – primers for unmethylated DNA sequence, HRM – high-resolution melting analysis, T_{ann} – annealing temperature

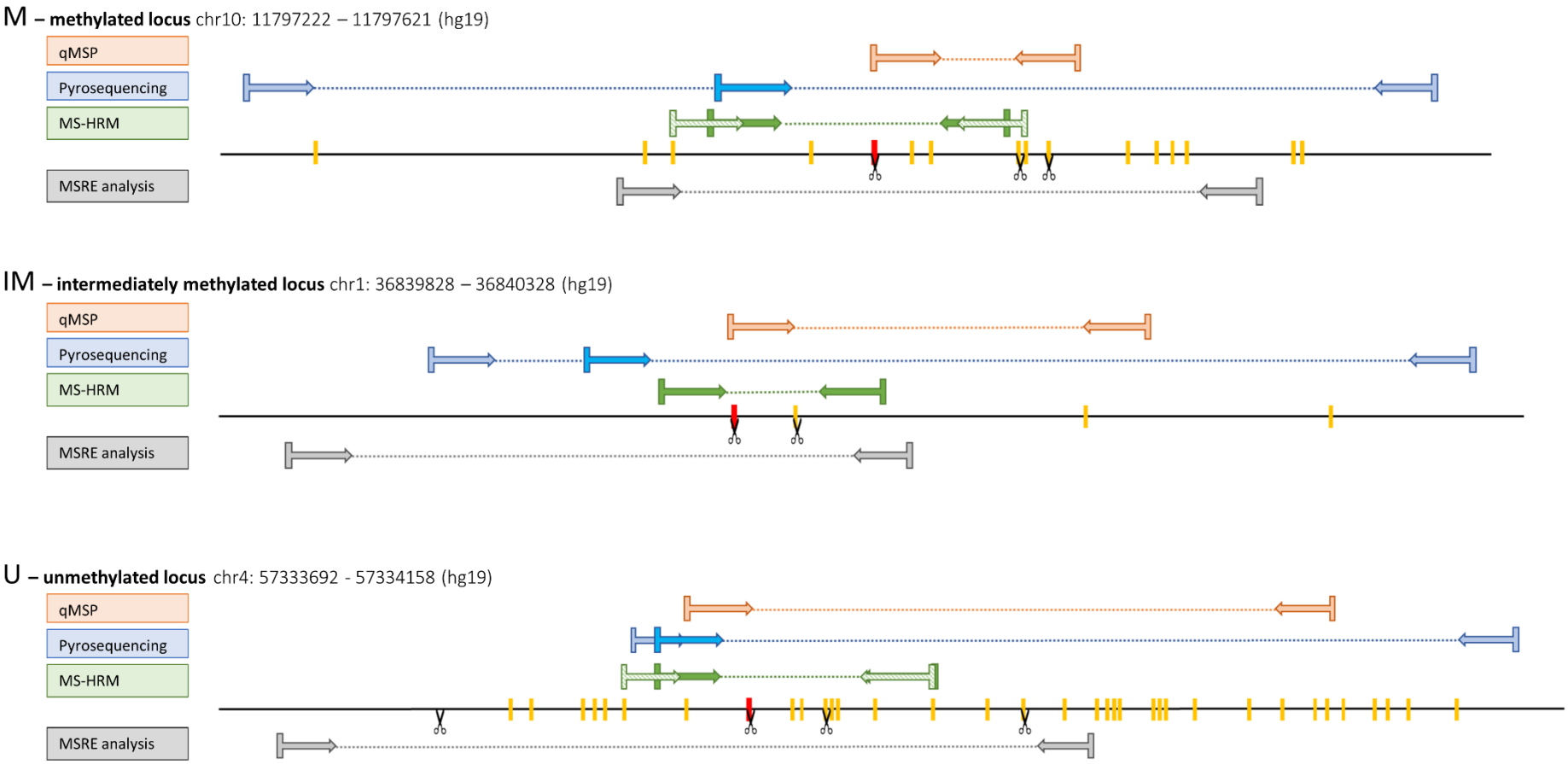


Figure 3 Positions of primer pairs, CpGs and restriction sites within studied regions. CpGs are shown as red and yellow bars on a line that represents the DNA sequence. Red CpG is the one chosen from Infinium MethylationEPIC BeadChip. Scissors indicate sites cleaved by MSRE. The patterned light green HRM primers were designed to include a CpG site on its 5'-end (M/U Wojdacz primers in **Table 9**). Lighter blue primers were used for initial pyrosequencing PCR, darker blue primers represent the sequencing primers.

3.2.6 Pyrosequencing

First, BS converted DNA (10–20 ng) was amplified using HotStar HiFidelity Polymerase Kit (Qiagen) with final $MgCl_2$ concentration adjusted to 2.5 mM. CoralLoad Concentrate (from PyroMark PCR kit, Qiagen) was added to final concentration of 1x for easy gel loading and to increase the specificity of primers. Final concentration of universal biotinylated and forward primers was 0.2 µM. Final concentration of reverse tailed primer was 0.04 µM. Recommended PCR reaction conditions for PyroMark PCR were used with 48 PCR cycles and T_{ann} according to **Table 9**. The quality of amplified PCR products was checked and the pyrosequencing was performed as described in chapter 3.1.5.

3.2.7 MS-HRM analysis

By mixing BS converted DNA of unmethylated and methylated standards, 100, 75, 50, 25, 10 and 0 % methylated standards were prepared. BS converted samples (15 ng) and standards (15 ng) were processed with EpiTect HRM PCR Kit (Qiagen). Final reaction volume was 20 µ with 0.375 µM final concentration of each primer. Reaction conditions were set according to the manufacturer's protocol. The experiment was performed using Rotor-Gene Q 2plex HRM Platform (Qiagen) with ramping for the HRM set from 67.1 to 82.2 °C, rising by 0.1 °C/2 s. Raw data were analyzed in web-based tool uAnalyze (Dwight *et al.*, 2012). First, a baseline normalization was performed and then difference curves for all standards and samples were computed with the 0 % methylated standard used as a reference curve. Next, we plotted the calibration curves in Microsoft Excel according to Tse *et al.* (2011) using either area under the curve (AUC) of the standards' processed data or their peak heights (Tse *et al.*, 2011). The methylation percentage of samples was calculated from these calibration curves.

3.2.8 Quantitative MS-PCR

BS converted DNA (10-15 ng) was measured via quantitative PCR with subsequent melting curve analysis. QuantiTect SYBR® Green PCR Kit (Qiagen) was used to prepare the reaction mix with 20 µl final volume and 0.5 µM final concentration of primers. Recommended cycling conditions (40 cycles) were kept with T_{ann} according to **Table 9**. All samples together with unmethylated and methylated DNA standards were amplified in three runs: with unmethylated methylation-specific primers (MSP Unm), methylated methylation-specific primers (MSP Met), and primers used for HRM analysis. The amplification efficiency was

computed for each primer set according to Pfaffl (2007). Quantitative PCR was performed with four 4-time dilutions of BS converted DNA for each sample and decadic logarithm of the dilutions was plotted against acquired threshold cycles (C_ts). The efficiency was calculated from the slope of the calibration curves according to the equation: $E = \left[10^{\left(-\frac{1}{slope}\right)} - 1\right] \cdot 100$. All measurements were performed on StepOnePlus instrument (Thermo Fisher Scientific). Methylation levels were subsequently calculated with all three methods reviewed in Husseiny *et al.* (2012).

3.3 DNA methylation sequencing panel

3.3.1 Patients

Together we analyzed 178 adult AML patients divided into a test cohort (n = 128, patients from Institute of Hematology and Blood Transfusion, characteristics are provided in **supplementary Table 2** in section 9) and a validation cohort (n = 50, patients from partner institution University Hospital Brno, characteristics are provided in **supplementary Table 3** in section 9). Patients were diagnosed between years 2013 - 2016 and treated with curative intent initiated with 3+7 induction regimen. Institutional Ethics Committees approved this study and all patients gave their full consent. The research conforms with The Code of Ethics of the World Medical Association.

3.3.2 DNA methylation sequencing panel

We designed a novel DNA methylation panel for targeted bisulfite sequencing. The panel comprised of 239 loci assigned to 186 genes. The custom probes were made by Roche company (Basel, Switzerland). The range of selected regions was 121–35,606 bp with average of 2,910 bp and median 1,473 bp. The total size of the panel was 573,406 bp. The investigated regions are listed in **supplementary Table 4** in section 9.

3.3.3 Samples' preparation and targeted bisulfite sequencing

Diagnostic DNA of AML patients was either isolated from whole peripheral blood with MagCore system (RBSBioscience) or from peripheral blood leukocytes lysed in RLT buffer using AllPrep DNA/RNA Mini Kit (Qiagen). We also analyzed 11 samples from healthy donors together with samples from the test cohort. Healthy donors' DNA was isolated from CD34+ cells harvested from buffy coats using MicroBeads kit (Miltenyi Biotec), the procedure is described also in chapter 3.1.2.

Each sequencing library consisted of 16-18 samples and libraries were prepared according to SeqCap Epi protocol (Roche) with KAPA HyperPrep Kit (Roche). First, Bisulfite-conversion Control (unmethylated DNA from Enterobacteria phage lambda, a component of SeqCap Epi Accessory kit (Roche)) was added to 800–1,200 ng of patients/donors' DNA. This DNA mixture was subsequently fragmented either via Bioruptor Pico instrument (Diagenode, Liège, Belgium) or E220 Focused ultrasonicator (Covaris, Woburn, MA, USA) to get ~ 200 bp DNA fragments. The size of fragments was checked in two random samples from each library on 4200 TapeStation System (Agilent Technologies, Santa Clara, CA, USA) using Agilent D1000 ScreenTape (Agilent Technologies). EZ DNA Methylation Lightning Kit (Zymo Research) was used for the bisulfite conversion, as recommended in the SeqCap Epi workflow. Each library was hybridized for ~ 68 hours with probes from our DNA methylation sequencing panel (chapter 3.3.2). KAPA Library Quantification Kit (Roche) was utilized to assess the final concentration of sequencing libraries. Average size of the libraries' fragments was measured on 4200 TapeStation System with Agilent D1000 ScreenTape (Agilent Technologies). Sequencing was performed using MiSeq Reagent Kit v2 (300-cycles) (Illumina) on MiSeq instrument (Illumina).

3.3.4 Sequencing data analysis

Raw sequencing data were acquired as fastq files. First, the quality of reads was checked using MultiQC (Ewels *et al.*, 2016) and FastQC (Andrews, 2010) software. Reads were subsequently trimmed and filtered with Cutadapt 2.4 (Martin, 2011) and the reads' quality was checked once more. Next, mapping tool Segemehl (Otto *et al.*, 2012) was used for mapping of the pre-processed data to human genome (version GRCh37/hg19) with added phage lambda sequence (NC_001416.1). The mapping statistic was computed and Samtools software was utilized for sorting and indexing of bam files that contained the mapped reads. Haarz tool (Otto *et al.*, 2012), with enabled "callmethyl" option, was subsequently used for the subtraction of methylated positions, creating vcf files. These files, containing all methylated positions, were further processed in R software. Positions that corresponded to the lambda phage sequence were separated and used for the bisulfite conversion ratio assessment for each sample. The other positions were filtered and only CpG positions were left in the data. Finally, data for all samples from each cohort (test and validation) were put together and only CpGs that were measured in more than 75 % of the samples were left in two summarizing tables.

3.3.5 MethScore computation

Initially, only the test cohort was analyzed. Cox univariate regression analysis for OS was performed for each detected CpG and only CpGs acquiring significant p-value (< 0.05) in this analysis were selected. Subsequently, using these CpGs, their methylation level and Cox regression coefficient beta, we computed a weighted summary score called MethScore. We adapted the procedure from Marcucci *et al.* (2014). MethScore (MS) for patient *i* was counted via linear combination as follows: $MS_i = \sum w_j \cdot x_{ij}$, where x_{ij} is a dichotomized methylation value (0 if methylation is < median methylation for the CpG in the whole cohort or 1 if methylation is ≥ median) for CpG *j* in patient *i* and w_j is the Cox regression coefficient for CpG *j*. The computation was performed in statistical software R. The same set of CpGs, with methylation significantly affecting the OS of patients from the test cohort, was used to assess the MethScore for the validation cohort.

3.3.6 Definitions and statistical analyses

Overall survival was set as time from diagnosis until death of any cause, event-free survival was set as time from the first complete remission until hematological relapse or death. Surviving patients were censored to the 6th April 2020. To evaluate the significance for OS and EFS, we used Kaplan-Meier curves with two-sided logrank test. Uni- and multivariate analyses were performed via Cox regression analysis. Data were corrected to the time of transplantation for the multivariate regression and the assumption of proportionality was checked for each regression model. We used Mann-Whitney test (Wilcoxon test) when comparing continuous variables and Fishers' exact test when comparing categorical variables in patients' groups. All statistical analyses were performed in R.

3.3.7 Gene ontology analysis

Free online tools were used. Bed files with positions of detected CpGs were submitted to GREAT (http://great.stanford.edu/public/html/) (McLean *et al.*, 2010) and Enrich (https://amp.pharm.mssm.edu/Enrichr/) (Kuleshov *et al.*, 2016). Enrich program generated gene lists from the submitted bed files, which we further submitted to GOrilla (http://cbl-gorilla.cs.technion.ac.il/) (Eden *et al.*, 2009).

4 RESULTS

4.1 DNA methylation and hydroxymethylation changes in AML patients with mutations in *DNMT3A*, *IDH1/2* or their combinations

4.1.1 DNA methylation profiles

Using the methylation arrays, we acquired data from 716,847 probes for each investigated sample (n = 28). With those, we performed hierarchical clustering analysis, the resulting dendrogram is shown in **Figure 4**. The analysis revealed three main clusters: cluster of *DNMT3A*^{mut} patients, *IDH1/2*^{mut} patients, and cluster with healthy donors' samples (controls). Majority of patients (63 %) with mutations in both *DNMT3A* and *IDH1/2* genes were included in the controls' cluster.

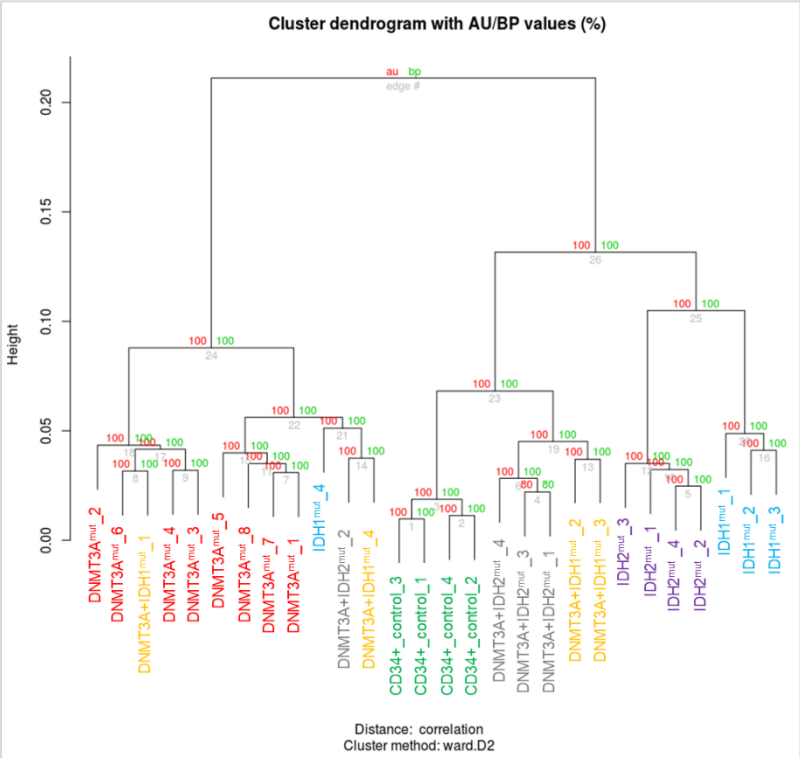


Figure 4 Hierarchical clustering of methylation profiles of AML samples (n = 24) and healthy controls' samples (n = 4) represented by oxBS β-values of 716,847 probes from Infinium MethylationEPIC BeadChip (Illumina).

Next, we were searching for differentially methylated positions (DMPs) between the groups of AML samples and controls. We considered the methylation difference (Δβ) as biologically significant for $-0.25 \leq \Delta\beta \geq 0.25$ with adjusted p-value < 0.05 as computed by the champ.DMP function in R software. The results are summarized in **Table 10**. An overall hypomethylation was observed in samples with mutated *DNMT3A*

and a hypermethylation tendency was found in samples with IDH1/2^{mut} samples. Noticeably fewer DMPs were detected in *DNMT3A&IDH1/2*^{mut} groups. Also, the ratio of hyper- and hypo-methylated sites was more balanced for *DNMT3A&IDH2*^{mut} samples, with a prevalent hypomethylation in *DNMT3A&IDH1*^{mut} group. A significantly more DMPs were found in the *IDH2*^{mut} group than in the *IDH1*^{mut} patients.

Table 10 Numbers of differentially methylated positions detected between AML sample groups and healthy donors' samples.

Sample group	Total number of DMPs	Hyper-DMPs	Hypo-DMPs	Common DMPs for specific groups	Common DMPs for *IDH1/2*^{mut} samples
DNMT3A^{mut}	32534	2420 (7 %)	30114 (93 %)	NA	NA
IDH1^{mut}	13830	8061 (58 %)	5769 (42 %)	9326 (5133 hyper, 4193 hypo)	218 (215 hyper, 3 hypo)
IDH2^{mut}	37349	24364 (65 %)	12985 (35 %)		
DNMT3A&IDH1^{mut}	6252	2269 (36 %)	3983 (64 %)	375 (353 hyper, 22 hypo)	
DNMT3A&IDH2^{mut}	7627	4022 (53 %)	3605 (47 %)		

DMPs – differentially methylated positions ($-0.25 \leq \Delta\beta \geq 0.25$; adjusted p-value < 0.05), Hyper – hypermethylated, Hypo – hypomethylated, NA – not analyzed

When we performed the gene ontology analysis, immune response and apoptosis pathways were associated with the 215 hypermethylated CpGs common for all *IDH1/2*^{mut} samples. Among the apoptosis-related genes, granzyme B (GZMB), a key protein for apoptosis induction (Chiusolo *et al.*, 2017), was found. Another interesting gene found in the gene ontology analysis was *CHFR* (checkpoint with forkhead-associated and RING finger domains), tumor suppressor gene that serves as a mitotic checkpoint and was recently described as a marker of poor prognosis when hypermethylated or downregulated (Gao *et al.*, 2016; Zhou *et al.*, 2018).

4.1.2 DNA hydroxymethylation profiles

The difference between measured BS and oxBS β-values corresponds to the hydroxymethylation level of each CpG from the array. We applied the champ.DMP function to find statistically significant differences ($\Delta\beta_{BS\text{-}oxBS}$ with adjusted p-value < 0.05) and the threshold for hydroxymethylation detection was set as 0.03335304 (threshold

computation is described in chapter 3.1.3). For the subsequent analyses, we selected 536,617 probes for which at least one sample passed this threshold (had $\Delta\beta_{BS\text{-}oxBS} > 0.03335304$).

In the hierarchical clustering analysis, dendrogram is shown in **Figure 5**, three clusters of specific groups emerged: controls, *IDH2*^{mut} patients and *DNMT3A*^{mut} patients. The last group of samples clustered together with healthy controls and comprised mostly of *IHD1*^{mut} and *DNMT3A&IDH1*^{mut} samples.

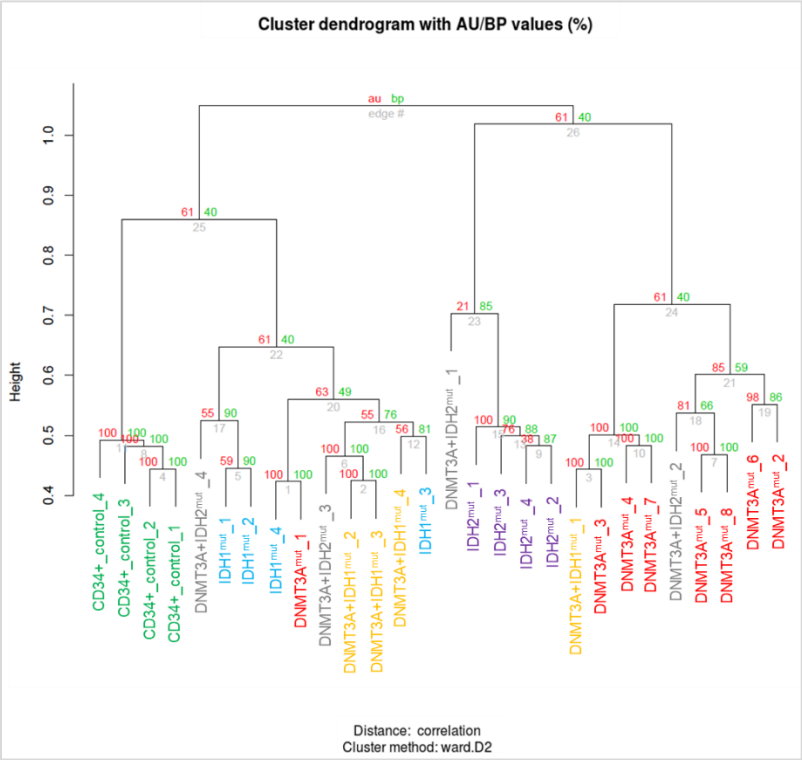


Figure 5 Hierarchical clustering of hydroxymethylation profiles of AML samples (n = 24) and healthy controls' samples (n = 4) represented by $\Delta\beta_{BS\text{-}oxBS}$ values of 536,617 probes. Probes with $\Delta\beta_{BS\text{-}oxBS} < 0.03335304$ were taken as not hydroxymethylated and assigned a value of 0.0001.

Next, we set a threshold of $\Delta\beta_{BS\text{-}oxBS} \geq 0.2$ for the selection of CpGs with biologically relevant DNA hydroxymethylation. In **Table 11**, we provide the numbers of detected hydroxymethylated positions (hmCpGs) in each group of samples. The numbers of hmCpGs were much lower when compared with the numbers of differentially methylated sites (**Table 10**). The lowest number of hmCpGs was found in the group of *DNMT3A*^{mut} patients and the highest number among *IDH2*^{mut} samples. Significantly more hydroxymethylated positions were detected in *IDH2*^{mut} samples than in *IDH1*^{mut} samples, a similar phenomenon as found when analyzing the methylation data (chapter 4.1.1).

Table 11 Hydroxymethylated CpGs detected in AML sample groups and healthy donors (controls)

Sample group	Total number of hmCpGs	Hyper-hmCpGs compared to controls	Unique hmCpGs for particular group	Common hmCpGs for *IDH1/2*^{mut} samples	Common hmCpGs compared to controls
controls	1268	NA	785 (62 %)	NA	NA
***DNMT3A*^{mut}**	133	66	11 (8 %)	NA	25
***IDH1*^{mut}**	715	478	150 (21 %)	50	
***IDH2*^{mut}**	1931	1595	1239 (64 %)		
***DNMT3A&IDH1*^{mut}**	709	400	143 (20 %)		
***DNMT3A&IDH2*^{mut}**	216	120	23 (11 %)		

hmCpGs – hydroxymethylated CpGs ($\Delta\beta_{BS\text{-}oxBS} \geq 0.2$, adjusted p-value < 0.05), Hyper-hmCpGs – hyperhydroxymethylated CpGs when compared to controls, NA – not analyzed

In the gene ontology analysis, a well-known protooncogene *MYB* was found to be hydroxymethylated and also upregulated (according to the results from the expression analysis, chapter 4.1.5) in *IDH2*^{mut} samples, see **Figure 6**. Another gene, *RNF2016* (ring finger protein 216), was found as hydroxymethylated in all samples, except the *DNMT3A*^{mut} group. This gene encodes a protein with fundamental role in cellular homeostasis. Upregulation of *RNF2016* was shown to be associated with progression of colorectal cancer (Wang *et al.*, 2016).

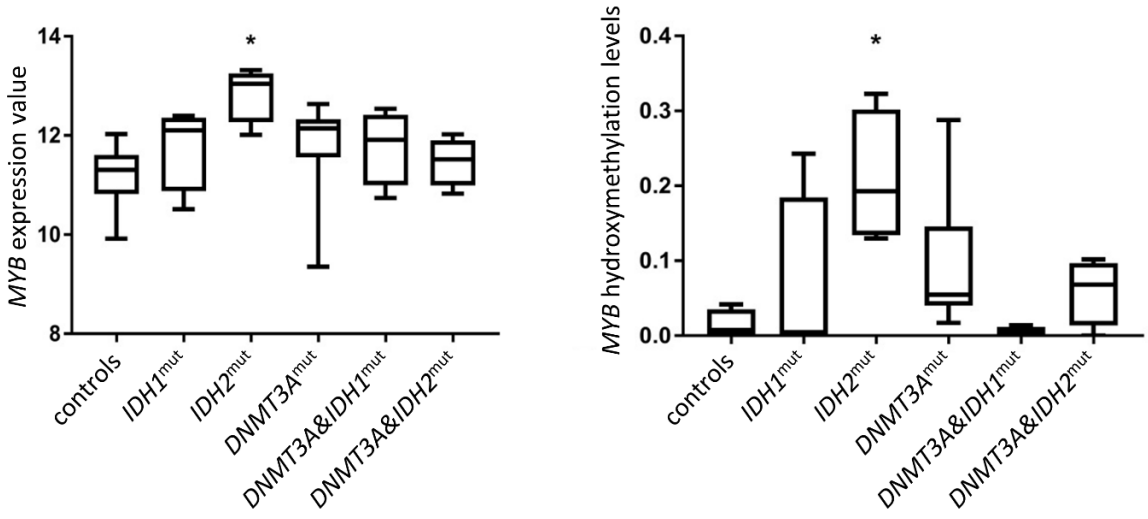


Figure 6 *MYB* expression and hydroxymethylation levels in all samples' groups. Expression levels were calculated as log2 normalized expression data from HumanHT-12 v4 Expression BeadChip (Illumina). Hydroxymethylatin data are represented by $\Delta\beta_{BS\text{-}oxBS}$, measured with Infinium MethylationEPIC BeadChip (Illumina). * p-value < 0.05 when compared with controls.

4.1.3 Validation of DNA methylation and hydroxymethylation data

We validated the methylation and hydroxymethylation levels measured via Infinium MethylationEPIC BeadChip with pyrosequencing. For the validation, we selected two methylated loci cg00771752 and cg00338702 (associated with *GZMB* and *CHFR* genes, respectively), and two hydroxymethylated positions cg23073132 and cg05617317 (associated with *MYB* and *RNF2016*).

The methylation levels were successfully validated in all samples (n = 28), with Pearson correlation coefficients R = 0.91 (p-value < 1.87e-5) for *GZMB* and R = 0.72 (p-value < 3,52e-7) for *CHFR* locus. However, we were not able to fully reproduce the hydroxymethylation data by pyrosequencing. For nine samples, we performed two pyrosequencing analyses, one with BS and one with oxBS converted DNA. Subsequently, we computed the difference $\Delta\beta_{BS\text{-}oxBS}$ but the values were significantly smaller (3.11 ± 3.82 for *RNF216* and 3.85 ± 3.21 for *MYB*) than those detected with arrays (20.15 ± 17.52 for *RNF216* and 9.79 ± 11.2 for *MYB*). Nevertheless, the correlation between original and validation data was quite good with R = 0.82 (p-value=0.002) for *RNF2016* and R = 0.70 (p-value= 0.024) for *MYB*.

4.1.3.1 DNA hydroxymethylation assessment with mass spectrometry

To confirm the assumption from previous two chapters, that DNA hydroxymethylation levels in peripheral blood samples are considerably lower than the levels of DNA methylation, we analyzed an independent cohort of AML samples at diagnosis (n = 40) and two healthy donors' samples with LC-MS/MS. There was a profound difference between the representation of 5-mC and 5-hmC in the samples, results are shown in **Figure 7**.

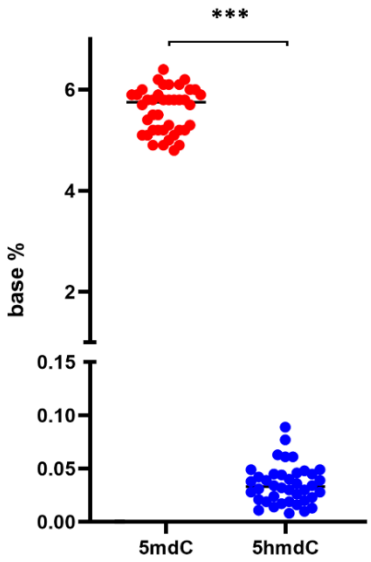


Figure 7 Levels of 5-deoxyhydroxymethylcytosine (5hmdC) and 5-deoxymethylcytosine (5mdC) in 40 AML patients and 2 healthy donors' samples measured with LC-MS/MS. Levels are expressed as percentage from total cytosine bases. *** p-value < 0.001.

4.1.4 Influence of co-mutations (other than *DNMT3A* and *IDH1/2*) on DNA methylation and hydroxymethylation profiles

Mutations in *NPM1* were the most often co-occurring mutations found in 15 AML samples (62.5 %) and it seems that the presence of *NPM1*^{mut} influenced the methylation profiles of investigated patients. *DNMT3A*^{mut} samples with mutated *NPM1* (*DNMT3A*^{mut}_1,3,4,6) and wild-type *NPM1* (*DNMT3A*^{mut}_2,5,7,8) had a tendency to cluster together. Also, only one sample out of the *IDH1/2*^{mut} groups (*IDH1*^{mut}_4) did not bear *NPM1*^{mut} and clustered away from the rest of IDH1/2^{mut} samples (**Figure 4**).

The presence of *FLT3*-ITD (in 10 out of 24 AML) had no obvious effect on DNA methylation profiles of the investigated samples.

There were two *DNMT3A*^{mut} patients (*DNMT3A*^{mut}_5,8) with co-occurring *TET2* mutations. These samples did not display any similarity with the DNA methylation profiles of *DNMT3A&IDH1/2*^{mut} samples and clustered within the rest of *DNMT3A*^{mut} samples (**Figure 4**). In the hydroxymethylation analysis, these two samples clustered together (**Figure 5**).

4.1.5 Overall expression profiles

After the initial data processing, we acquired 45,607 probes with which we performed differential gene expression analysis, results are summarized in **Table 12**. We compared each group of patients with healthy donors' samples. The highest numbers of deregulated genes were detected among *DNMT3A*^{mut} and *IDH2*^{mut} patients. On the contrary, the group with the least different expression profile was *DNMT3A&IDH2*^{mut}. In the gene ontology analysis,

most of the differentially expressed genes were linked to DNA binding, transcription, splicing, and translation processes.

To compare the overall gene expression profiles of investigated samples, we performed a hierarchical clustering analysis with probes that had detection p-value < 0.05 (n = 10,067), resulting dendrogram is shown in **Figure 8**. The cluster did not reproduce the results obtained from either DNA methylation or hydroxymethylation analyses (**Figures 4** and **5**). Only the group of healthy donors (controls) formed a united cluster.

Table 12 Summary of differential gene expression analysis between healthy donors and groups of AML patients

AML sample group	Total number of DEG	Gene ontology	Number of up/down regulated DEG		Gene ontology
DNMT3A ^{mut}	826	immunity, hematopoiesis, signal transduction, RNA, translation through ER, DNA binding	up	294	mRNA splicing, neutrophils deregulation
			down	532	SRP transport, viral transcription
IDH1 ^{mut}	319	splicing, transcription, dendritic cell differentiation, mRNA binding, kidney/liver cancer	up	145	dendritic cell differentiation
			down	174	splicing
IDH2 ^{mut}	730	DNA binding, viral transcription, SRP transport, cancer, diabetes	up	322	viral transcription, SRP transport, cancer
			down	398	splicing
DNMT3A& IDH1 ^{mut}	361	transcription regulation, viral transcription, DNA binding, Ser/Thr kinase activity, RNA binding, kidney cancer, eosinophilia	up	147	transcription regulation, SRP transport, kidney/liver cancer, carcinoma
			down	214	regulation of transcription, Golgi to membrane transport
DNMT3A& IDH2 ^{mut}	161	transcription, translation, transporters activity, liver cancer	up	30	transcription, miRNA processing, liver/kidney cancer, melanoma
			down	131	translation, cancer, eosinophilia

DEG – differentially expressed genes, ER – endoplasmic reticulum, SRP – signal recognition particle

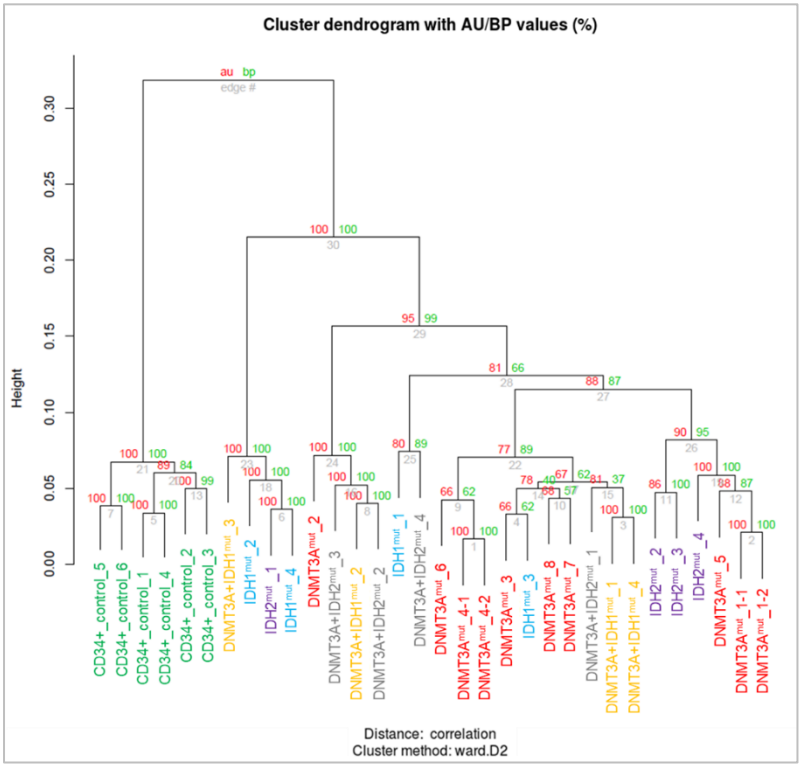


Figure 8 Hierarchical clustering analysis of gene expression data measured with HumanHT-12 v4 Expression BeadChip (Illumina). Investigated samples, AML (n = 24) and healthy controls (n = 4), are represented by 10,067 probes (with detection p-value < 0.05). Two technical replicates were analyzed with two different arrays (DNMT3A^{mut}1-1/2; DNMT3A^{mut}4-1/2).

4.1.6 *CHFR* methylation levels as a potential prognostic biomarker

When we performed the gene ontology analysis of aberrantly methylated CpGs, a tumor suppressor *CHFR* was detected among the differentially methylated sites in all *IDH1/2*^{mut} samples (chapter 4.1.1). Therefore, we measured the DNA methylation (at cg00338702) and *CHFR* expression in a statistically significant number of patients (n = 104) to see if there is a connection.

Only 4 % of investigated AML samples exhibited downregulated expression (values < average – 2·SD, standard deviation). These samples also showed a significantly higher methylation in associated locus (average 92.8 ± 6.6 %) in comparison with the rest of the samples (19.2 ± 15.0 %, p-value < 0.0001). Due to the low number of patients with hypermethylation and downregulated *CHFR* expression, we were not able to analyze the prognostic significance of this aberration.

4.1.7 *GZMB* methylation levels as a potential prognostic biomarker

Similar to *CHFR*, aberrant methylation of apoptosis-related gene *GZMB* was also detected in all *IDH1/2*^{mut} samples (see chapter 4.1.1). The differentially methylated region is located

approximately 40 kbp upstream of the gene and covers two adjacent CpGs. We examined the site via pyrosequencing in 104 AML patients and 10 healthy donors. Based on the DNA methylation levels detected in the cohort of healthy donors, we set an upper methylation threshold of 45 %. Subsequently, we divided the AML patients into three groups: unmethylated (UM) with methylation level < 45 % at both CpGs (n = 40); intermediately methylated (IM) with one CpG > 45 % and one CpG < 45 % methylated (n = 15); and hypermethylated (HM) with both CpGs > 45 % methylated (n = 49).

When we investigated the gene expression of *GZMB* in these groups, we found only a weak correlation with methylation levels for UM (R = 0.3, p-value = 0.12) and HM (R = -0.4, p-value = 0.02) patients. However, we discovered a significant difference in survival between these groups with *GZMB* hypermethylation being a marker of inferior prognosis (see **Figure 9**).

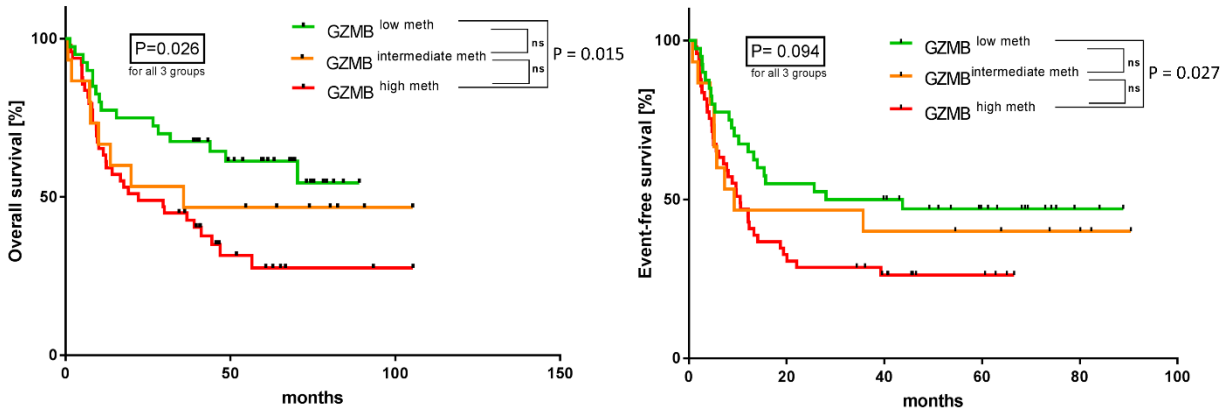


Figure 9 Kaplan-Meier curves for overall and event-free survival of AML patients divided by methylation status of *GZMB*-associated region.

We confirmed the prognostic implication of aberrant *GZMB* methylation in a Cox multivariate analysis, where it remained significant for overall survival of patients, together with age, presence of *FLT3*-ITD, and transplantation in the first remission (**Table 13**). *GZMB* hypermethylation was not among the prognostically significant covariates for event-free survival, where only ELN classification (Döhner *et al.*, 2017), *FLT3*-ITD, and transplantation remained statistically significant.

Table 13 Multivariate Cox regression analysis for overall and event-free survival of AML patients.

Overall survival				
Variable	**HR**	**95% CI**	**p-value**	
GZMB methylation	1.446	1.027-2.036	**0.035**	*
Prognostic group according to ELN	1.612	0.952-2.729	0.076	ns
Age at diagnosis	1.036	1.008-1.065	**0.012**	*
Leukocytes at diagnosis	1.000	0.995-1.005	0.933	ns
Percentage of blasts in peripheral blood	0.993	0.982-1.003	0.169	ns
FLT3-ITD	3.257	1.381-7.679	**0.007**	**
NPM1 mutation	0.652	0.317-1.339	0.244	ns
Transplantation in the 1st CR	0.378	0.191-0.748	**0.005**	**
Event-free survival				
Variable	**HR**	**95% CI**	**p-value**	
GZMB methylation	1.202	0.891-1.622	0.229	ns
Prognostic group according to ELN	2.465	1.501-4.048	**0.0001**	***
Age at diagnosis	1.013	0.991-1.036	0.246	ns
Leukocytes at diagnosis	1.001	0.996-1.006	0.604	ns
Percentage of blasts in peripheral blood	1.000	0.991-1.010	0.960	ns
FLT3-ITD	2.696	1.273-5.711	**0.01**	*
NPM1 mutation	0.737	0.385-1.411	0.356	ns
Transplantation in the 1st CR	0.178	0.088-0.360	**0.0001**	***

HR- hazard ratio, CI – confidence interval, ELN – European LeukemiaNet classification (Döhner *et al.*, 2017), CR – complete remission, ns – not significant

Because we did not find a reliable connection between *GZMB* methylation and expression, we tested whether the negative prognostic value could be explained by the presence of *IDH1/2* mutations (the aberrant hypermethylation of GZMB gene was discovered in *IDH1/2*^{mut} samples). But only 35 % of patients from the HM group had mutation in either *IDH1* or *IDH2* gene.

We also tested whether the *GZMB* hypermethylation is linked only to the leukemic cell population and thus dependent on the percentage of blasts at diagnosis. We selected 23 patients from the original cohort (n = 104) and measured the DNA methylation of *GZMB* associated region in sorted blast populations (CD34+/CD117+ cells). The methylation levels for both positions correlated well (R = 0.95 with p-value = 1.45e-11 and R = 0.91 with p-value = 5.33e-9) and samples belonged to the same UM/IM/HM category as if divided by methylation measured on whole-blood DNA.

4.2 DNA methylation validation methods

4.2.1 MSRE analysis

With MSRE analysis, we were able to accurately measure the DNA methylation levels at methylated (M) and unmethylated (U) regions. The average values for all investigated samples (n = 10) were 95.3 ± 1.8 % and 2.3 ± 0.5 % for M and U regions, respectively. However, for the intermediately methylated (IM) region, we obtained lower methylation levels than expected, the average was 12.6 ± 2.2 %. Thus, we tried to shorten the digestion time in order to see if it will increase the methods' accuracy for the IM region. We performed two additional experiments with 1.5-hour and 1-hour digestions (a 2-hour digestion was recommended in the original protocol). We measured four samples and there was only a slight change in the acquired methylation levels for IM region after shorter cleavage to 13.3 % (1.5-hour digestion) and 17.0 % (1-hour digestion), see **Figure 10**. The results for M and U regions remained unchanged.

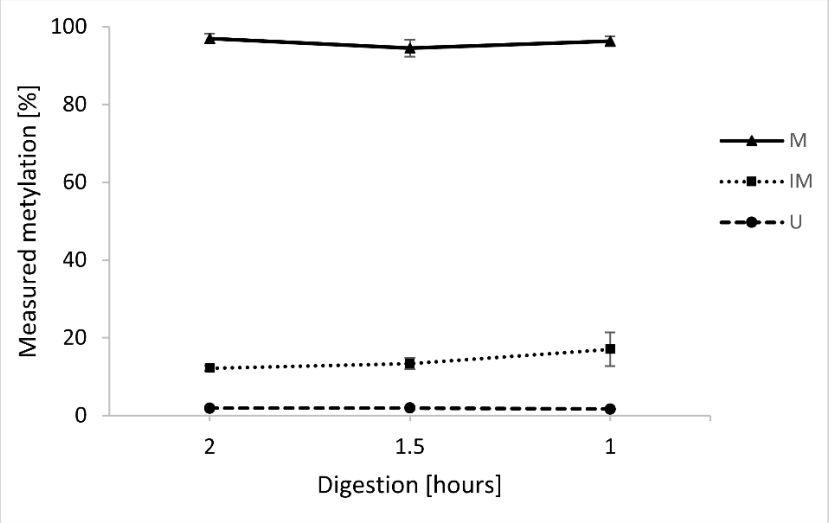


Figure 10 Methylation levels measured with MSRE analysis with various digestion times. Error bars represent the standard deviation for four samples. M – methylated region, IM – intermediately methylated region, U – unmethylated region.

4.2.2 Pyrosequencing

The most important step in pyrosequencing is the initial PCR where a strong amplicon must be obtained. We always checked the quality of amplicons on agarose gel electrophoresis, see **Figure 11**.

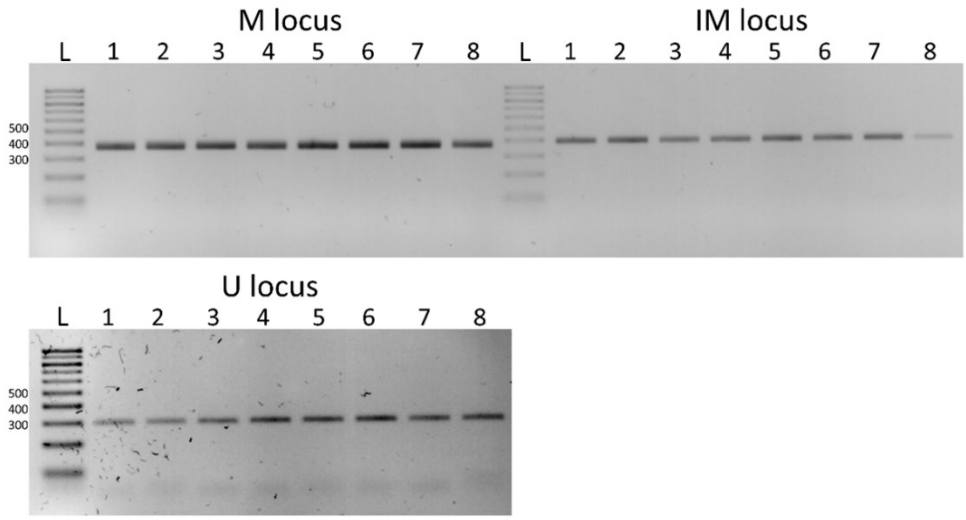


Figure 11 Agarose gel electrophoresis of eight samples (marked 1-8 in the picture) after the initial PCR for pyrosequencing of methylated (M), intermediately methylated (IM) and unmethylated (U) loci. L - 100 bp DNA ladder, 300-500 bp bands are marked in the picture.

However, even after we acquired satisfactory amplicons, we did not achieve a sufficient signal during pyrosequencing. Thus, we decided to enhance the binding of the PCR product to the streptavidin beads by adding gradually 1, 2 and 3 µl of the beads per sample. Samples where 2 µl of streptavidin beads were added gained the strongest signal on pyrogram. To further increase the number of bound PCR amplicons, we prolonged the agitation step to 20 minutes. It was also essential to proceed immediately to the next step after the agitation, because the beads must be well resuspended in the tube in order to be efficiently taken up by the probes, used in the subsequent step. After these adjustments, the peak heights in resulting pyrograms varied from 50 to 200 units, which was satisfactory since at least 40 units are required according to the manufacturer's recommendations for proper measurements.

In our experiment, we measured the DNA methylation of four regions that included selected M, IM and U loci in ten samples. In the region around M locus, all other CpG sites (n = 3) were methylated, average methylation for the whole region was 95.4 ± 3.1 %. We obtained similar result for the region surrounding U locus, where all investigated CpGs (n = 3) were unmethylated with average methylation 7.4 ± 3.1 %. There were two CpGs in the intermediately methylated region. The average methylation level of the selected IM CpG was 58.5 ± 7.3 %, however, the other CpG was rather unmethylated with average 18.4 ± 3.8 %, resulting in the overall methylation for the whole investigated IM region of 37.7 ± 21.4 %.

4.2.3 MS-HRM

For the MS-HRM experiments, we designed two sets of primers for M and U regions. One primer pair did not have any CpG in in their sequence and the other pair had one or two CpGs in its 5'-end (called Wojdacz primers, see **Table 9**). For the IM region, we only had primers without CpGs because of the lack of CG dinucleotides in the area (see **Figure 3**).

From measured MS-HRM data, we computed calibration curves for each primer set as described by Tse *et al.* (2007). In **Table 14**, a summary of calculated methylation levels for each region together with coefficients of determination (R^2) for the peak-height based calibration curves is provided. We also computed the methylation levels from AUC-based calibration curves, see **Table 15**. In both cases, the R^2 coefficients were somewhat lower for M region and, on the contrary, slightly better for U region when Wojdacz primers were used.

Table 14 Coefficients of determination and counted methylation levels for peak height-based MS-HRM calibration

Region name	Primer set	R^2	Average methylation [%] (n = 10)	± SD
M	M HRM	0.95	93.61	5.28
M	M HRM Wojdacz	0.80	85.49	5.13
IM	IM HRM	0.97	29.20	4.71
U	U HRM	0.87	2.69	1.04
U	U HRM Wojdacz	0.94	0.57	0.81

M - methylated, IM – intermediately methylated, U – unmethylated, R^2 – coefficient of determination, SD – standard deviation

Table 15 Coefficients of determination and counted methylation levels for AUC-based MS-HRM calibration

Region name	Primer set	R^2	Average methylation [%] (n = 10)	± SD
M	M HRM	0.96	95.43	5.85
M	M HRM Wojdacz	0.85	148.58	8.68
IM	IM HRM	0.93	16.73	3.72
U	U HRM	0.85	1.95	0.69
U	U HRM Wojdacz	0.95	4.88	7.82

M - methylated, IM – intermediately methylated, U – unmethylated, R^2 – coefficient of determination, SD – standard deviation

With Wojdacz primers, lower methylation levels were obtained for both M and U regions when using the peak-height calibration and higher methylation levels when employing the AUC-based calibration. Furthermore, the methylation of M region measured with Wojdacz primers is disproportionately high when using the AUC-based calibration. The methylation levels acquired with primers without any CpGs are similar for both

calibration approaches. The methylation measured for IM region was more accurate with the peak-height based calibration.

4.2.4 Quantitative MS-PCR

To perform the qMSP experiments, we designed two sets of primers for each region, methylated (Met) and unmethylated (Unm). In all investigated samples (n = 10), the M region was amplified only by Met primers, the U region only by Unm primers, and IM region by both sets of primers. At the same time, the DNA standards were always amplified only with the corresponding primer set, see **Figure 12.**

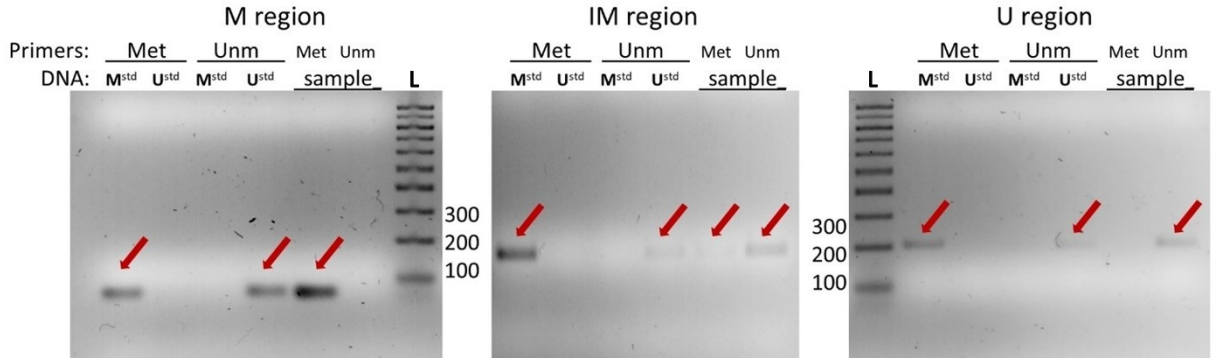


Figure 12 Specificity test of primers designed for qMSP experiments. Methylated DNA standard (M^{std}), unmethylated DNA standard (U^{std}), and one sample were amplified. Each region (M-methylated, IM – intermediately methylated, U - unmethylated) was amplified with primers specific for methylated DNA sequence (Met) and unmethylated DNA sequence (Unm). Agarose gel electrophoresis was performed with 100 bp DNA ladder (L), sizes 100 - 300 bp are marked. Red arrows point to visible amplicons.

We used a quantitative approach to evaluate the methylation levels of investigated samples. For these purposes, we used the HRM primers as methylation-unspecific primers. In **Table 16**, a summary of C_t values and amplification efficiencies for each primer set are provided. We were not able to properly measure the efficiency for IM Unm primers due to a high deviation between duplicates. The amplification of the first dilution for this primer pair started always after the 34th cycle, so we assume that the efficiency was quite low. Thus, despite the great coefficient of determination for the calibration curve ($R^2 > 0.99$), the results were not entirely reliable. Therefore, we chose a different approach to count the primers' efficiency based on standards' C_t values and an assumption that HRM primers have 100 % amplification efficiency. With following equation $E_c = 100 \cdot \frac{C_t^{BSP}}{C_t^{MSP}}$, we computed the counted efficiency for each primer pair and it corresponded well with measured

efficiencies assessed with the classical approach (Pfaffl, 2007), see **Table 16**. For the IM Unm primers, the counted efficiency was used in the subsequent analysis.

Table 16 Threshold cycles and primer pairs' amplification efficiencies in qMSP experiments

Region and used primer set	Average C_t of samples (n=10)				C_t of standards		Amplification efficiency		
	MSP	± SD	HRM	± SD	MSP	HRM	MSP measured	MSP counted	HRM measured
M Met	23.18	0.52	22.88	0.27	24.01	22.7	96.47	94.57	83
IM Met	34.54	1.23	24.33	0.35	29.23	24.51	81.13	83.85	94.78
IM Unm	32.21	0.5			31.99	25.15	125.8	78.63	
U Unm	37.36	0.8	23.04	0.2	37.95	23.4	65.93	61.66	90.52

M - methylated, IM – intermediately methylated, U – unmethylated, Met – primers for methylated DNA sequence, Unm – primers for unmethylated DNA sequence, MSP – methylation specific primers, HRM –methylation independent primers, SD – standard deviation

Data measured with the quantitative PCR were analyzed with three approaches, reviewed by Husseiny *et al.* (2012). Demethylation index and $\Delta\Delta C_t$ method provided similar results. The relative expression ratio gave highly variable results with extreme standard deviations. Results for all three methods are shown in **Table 17**.

Table 17 Summary of qMSP methylation results calculated using three different approaches

Region and used primer set	Demethylation index		$\Delta\Delta C_t$		Relative expression ratio	
	Average (n=10)	± SD	Average (n=10)	± SD	Average (n=10)	± SD
M Met	2.02	0.60	2.09	0.66	204.46	256.01
IM Met	0.05	0.03	0.03	0.02	1.92E-09	3.63E-09
IM Unm	0.51	0.10	0.50	0.12	0.02	0.02
U Unm	1.03	0.30	1.30	0.66	8.77	13.27

M - methylated, IM – intermediately methylated, U – unmethylated, Met – primers for methylated DNA sequence, Unm – primers for unmethylated DNA sequence, SD – standard deviation

The M region was highly amplified by the Met primers, the amplification was seemingly double in comparison with HRM primers (indicated by demethylation index/$\Delta\Delta C_t \sim 2$). The fact, that amplification efficiency of Met primers was higher than HRM primers may explain this discrepancy. We also tried to repeat the experiment with five samples and the $\Delta\Delta C_t$ was 1.5 ± 0.3 for the M region and Met primers, so another reason for the inconsistency may be simply the method's inaccuracy. The results for U region were close to 1, meaning a comparable amplification by HRM and U Unm primers. Regarding the IM region, the IM Unm primers amplified about half of the molecules in comparison with HRM primers, which would be an expected result. However, there was only little relative

amplification of the region by IM Met primers. The discrepancy in the computation may be caused by the great difference between C_t for methylated DNA standard (29.23) and samples (34.54) resulting in a low ratio of molecules amplified by IM Met primers. This was probably due to the higher affinity of IM Met primers to methylated DNA standard in comparison with samples, where the IM region is rather unmethylated with one CpG ~ 58 % methylated and the other ~ 18 % methylated (as measured by pyrosequencing, chapter 4.2.2).

4.2.5 Overall comparison of investigated methods

We compared the methylation results acquired by each method, see **Figure 13**. All methods correlated well with $R^2 > 0.92$ and p-value of regression analysis < 1.2e-17, except the qMSP results that were spoiled by very high standard deviations. Furthermore, we evaluated the investigated methods with regard to financial requirements (**Table 18**) and a few other perspectives (**Table 19**).

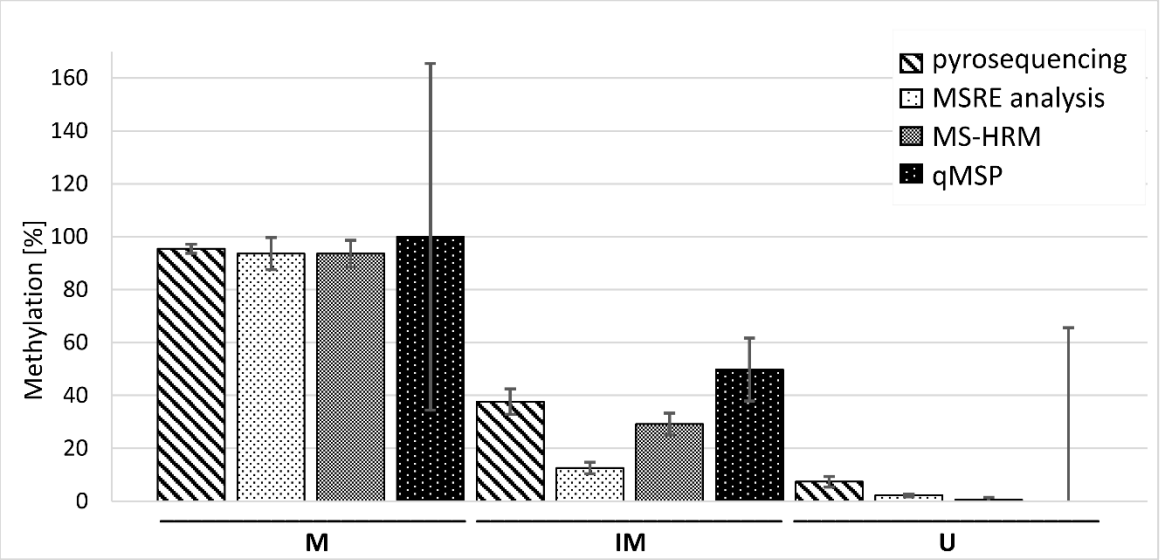


Figure 13 Average DNA methylation of ten samples measured with four different methods. Error bars represent the standard deviation. MSRE data were measured after 2-hour digestion. MS-HRM data were acquired using HRM M, HRM IM and HRM U Wojdacz primers. qMSP data were calculated using $\Delta\Delta C_t$ approach and multiplied by 100 to gain percentage. For the M locus, all values were higher than 100 % so we set the mean to 100 % to make the figure more comprehensible. For the IM and U loci in qMSP, we calculated the methylation percentage as 1-(Unm ΔΔCt). The qMSP SDs were calculated from the original values multiplied by 100. M – methylated region/locus, IM – intermediately methylated region/locus, U – unmethylated region/locus.

Table 18 Summary of costs for each method

Method	Total cost of analysis [$]	Number of samples measured	Number of standards measured	Total number of measurements	Cost per measurement [$]
MSRE analysis	576	10 for each locus[1], Test and Reference reaction, duplicates	2 for each locus[1], duplicates	144	4
pyrosequencing	162[3]	10 for each locus[1]	2 for each locus[1]	36	4.5
MS-HRM	85	10 for each locus[1], duplicates	6 for each locus[1], duplicates	96	0.9
qMSP	196	10 for each primer set[1], duplicates	2 for each primer set[2], duplicates	216	0.9

[1]Number of loci = 3
[2]Number of MSP/HRM primer sets for each locus = 3
[3]price of the pyrosequencing instrument ca 45 000 $

Table 19 Overall evaluation of investigated methods

Method	Base resolution	Consistency across methylation levels	Analysis of acquired data	Method optimization	Time consumption	Price
MSRE analysis	-	*	*	*	*	***
Pyrosequencing	+	***	*	*	***	**
MS-HRM	-	***	**	*/**(if needed)	*	*
qMSP	-	**	***	***	**	**

* - simple/low, *** - demanding/high

Pyrosequencing has most advantages when assessing DNA methylation of a single locus thanks to its base resolution. The primer design and results' interpretation are also very straightforward with provided software. Optimization is required only for the initial PCR step but is not always necessary. The only disadvantage is the high price of the pyrosequencing instrument. The price per measured sample is higher too because the method comprises of three steps: PCR, gel electrophoresis, and sequencing.

When the pyrosequencing is financially unaffordable, MS-HRM seems to be the second-best choice. Primer design is feasible for most regions, in our case we designed primers for CpG rich (M and U regions) as well as CpG poor (IM region) DNA sequences without problems. Thorough optimization of primer sequences and annealing temperatures is not necessary. The method itself is very simple, time and cost effective. Approximate methylation levels can be derived right from the melting curves. Exact quantification is a bit challenging without a specific chargeable software. However, it is feasible with a free online software and Microsoft Excel, as we showed in our analysis.

MSRE analysis is a quick and simple method with easy data assessment. The main advantage is that it does not require BS conversion and thus less DNA amount is needed and also the primer design is easier. The downside of this method is the higher price per measurement.

The last evaluated method, qMSP, was the most challenging one. The primer design was very difficult and nearly impossible for the CpG poor IM region. Finding a suitable annealing temperature at which both Met and Unm primers were specific only for the appropriate methylated or unmethylated allele of the region but at the same time amplified the corresponding DNA standard was very exacting. The performance of the measurement is also a bit demanding because it requires the amplification of a chosen region by at least one MSP primer set and one methylation unspecific primer set. The exact quantification of the methylation levels was not easy, too.

4.3 DNA methylation sequencing panel

4.3.1 Acquired sequencing data and their comparison with source literature

Two independent groups of patients, test (n = 128) and validation (n = 50) cohorts, were sequenced. For each sample, the bisulfite conversion ratio was > 99 % and more than 80 % of all reads were properly mapped, therefore, our data met the quality requirements

and could be used in further analyses. We obtained the methylation levels of 48,128 CpGs for the test cohort and 56,246 CpGs for the validation cohort.

Our sequencing panel was largely based on previously published studies (from years 2011 - 2019) that evaluated prognostic significance of certain DNA methylation changes. We wanted to validate the reported effect of DNA methylation on patients' survival with the data acquired for the test cohort. We selected regions that corresponded to the original publications and tried to use the same thresholds for DNA methylation levels. In cases where it was not possible due to a larger difference between published and our DNA methylation levels, we set the threshold as median or average methylation for all AML patients (whatever was closer to the original study). This was mainly a case when authors used qMSP for DNA methylation assessment. In our comparison, we were able to confirm the prognostic significance of DNA methylation in six regions out of nineteen investigated, a summary of the results is presented in **Table 20**.

The regions where we were able to confirm the predictive effect of DNA methylation were published in five studies. In three of them, authors measured the DNA methylation with microarrays (Deneberg *et al.*, 2011; Božić *et al.*, 2015; Qu *et al.*, 2017), NGS was used in one publication (Marcucci *et al.*, 2014) and the last one utilized bisulfite sequencing (Lin *et al.*, 2011). One study investigated the methylation of only one locus (Božić *et al.*, 2015) and in two studies only one (Lin *et al.*, 2011) or two (Qu *et al.*, 2017) small regions were examined. Two studies based their prognostication on a summarizing methylation score computed from a number of regions, mostly CGIs (Deneberg *et al.*, 2011; Marcucci *et al.*, 2014).

Table 20 Validation of prognostic significance of previously published DNA methylation changes.

Publication	Methylation detection method	Gene/region tested	Methylation threshold	P-value		Significance[1]
				Logrank (Mantel-Cox) test	Gehan-Breslow-Wilcoxon test	
Deneberg *et al.* (2011)	HumanMethylation27 BeadChip	*BHMT, CHST13, COL21A1, CST11, ESM1, ETNK2, GAL3ST3, KCNA1, KCNA6, KCNK4, MEGF10, MGC39715, NEUROG1, OXGR1, SLC4A11, SNCAIP, TCF15, TF*	30 % (AML average methylation of all genes)	0.025	0.032	*
Lin *et al.* (2011)	Bisulfite sequencing, quantitative MassArray	*CEBPA* promoter	5 % (AML[2] median)	0.002	0.001	**
Hájková *et al.* (2012)	MethyLight	*CDKN2B, ESR1, MYOD1, CALCA, SOCS1, CDH1*	cumulative methylation value[3] < 7	0.475	0.433	ns
		SOCS1 promoter	0.2 % (AML median)	0.472	0.614	ns
Treppendahl *et al.* (2012)	Bisulfite sequencing, pyrosequencing, MS-HRM	*VTRNA2-1* promoter	10 %, 38 %	0.871	0.892	ns
Hájková *et al.* (2014)	NGS, pyrosequencing	*PBX3* (*TAF1* binding site)	20 % (average AML)	0.319[4]	0.122[4]	ns
Jost *et al.* (2014)	TCGA data, pyrosequencing (validation)	1 CpG in *DNMT3A* promoter[5]	0 %	0.999	0.815	ns
Marcucci *et al.* (2014)	NGS: MethylCap enriched by MBD2, validation by RRBS and MassArray	*CD34, RHOC, SCRN1, F2RL1, FAM92A1, MIR155HG, VWA8*	low methylation (> 4 out of 7 genes have lower methylation than average in healthy controls)	0.004	0.004	**
Božić *et al.* (2015)	TCGA data, pyrosequencing (validation)	1 CpG in *C1R* 5'UTR region	27 %	0.031	0.019	*

Publication	Methylation detection method	Gene/region tested	Methylation threshold	Logrank (Mantel-Cox) test	Gehan-Breslow-Wilcoxon test	Significance[1]
Zhou *et al.* (2016)	qMSP, validation by bisulfite sequencing	*DLX4*	20 % (AML average)	0.128	0.049	ns
Guo *et al.* (2017)	qMSP	*SFRP1* promoter	15 % (AML average)	0.830	0.920	ns
		SFRP2 promoter	15 % (AML average)	0.236	0.196	ns
		SFRP1, SFRP2	14,7 % (AML average)	0.186	0.175	ns
Kelly *et al.* (2017)	DREAM, pyrosequencing, validation on TCGA data	*NPM2, SCGB3A1, CDKN2B, OSCP1*	20 % (AML average methylation of all genes)	0.230	0.217	ns
Li *et al.* (2017)	qMSP, Bisulfite sequencing	*NKD2* promoter	5 % (AML median)	0.131	0.082	ns
Liu *et al.* (2017)	qMSP	*RASSF1* promoter	1 %	0.342	0.196	ns
Qu *et al.* (2017)	CHARMcox, pyrosequencing (validation)	*UZTS2*	40 % (AML median)	0.090	0.062	.
		NR6A1	15 % (AML median)	0.015	0.009	*
		UTZS2, NR6A1	methylation < median methylation level in both genes	0.031	0.017	*
Zhou *et al.* (2017)	qMSP	*GPX3*	2 % (AML average)	0.227	0.100	ns
Šestáková *et al.* (2019)	MethylationEPIC BeadChip, pyrosequencing (validation)	2 CpGs in *GZMB* associated IGR	45 % at both/one/none of the two CpGs	0.781	0.897	ns

[1] p-value for both logrank and Wilcoxon tests < 0.1 (.); <0.05 (*); <0.01 (**)
[2] Excluded patients with favorable cytogenetic profile, $NPM1^{mut}$ a $CEBPA^{mut}$
[3] Cumulative methylation value = (1·number of genes with methylation < 15 %) + (2·number of genes with methylation 15-50 %) + (3·number of genes with methylation > 50 %); range 4-16 for our data
[4] Data for cumulative incidence of relapse
[5] Data for 2 surrounding CpGs, the particular one was not sequenced
MassArray - Mass spectrometry analysis of cleaved fragments of chosen regions amplified by PCR, TCGA data - data from The Cancer Genome Atlas Research Network 2013 AML study (Ley *et al.*, 2013), DREAM - Digital Restriction Enzyme Analysis of Methylation, qMSP - quantitative methylation-specific polymerase chain reaction, CHARMcox - Comprehensive High-throughput Array-based Relative Methylation Analysis combined with Cox proportional Hazards Model, RRBS - Reduced representation bisulfite sequencing, MS-HRM - methylation specific high-resolution melting analysis

4.3.2 MethScore

We developed a complex approach for prognostic evaluation of all loci targeted by our sequencing panel. First, we estimated the effect of DNA methylation level of each detected CpG on OS of patients from the test cohort using univariate Cox regression. We subsequently selected only those CpGs significant in the analysis (with $p < 0.05$) and got a set of 1,961 CpGs. Positions and average methylation levels of these CpGs are shown in **Figures 14** and **15**.

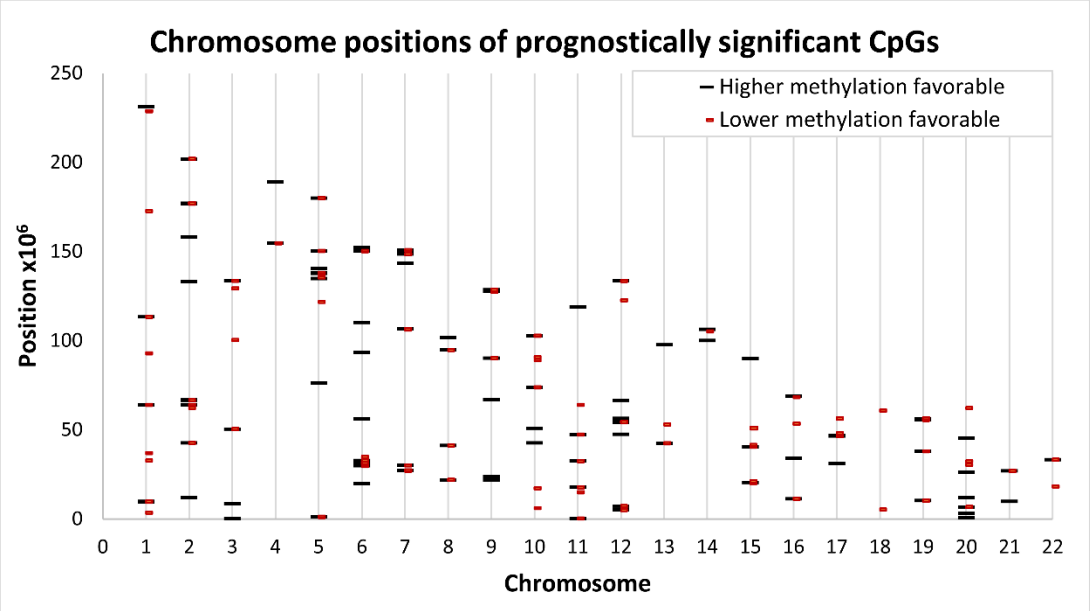


Figure 14 Positions of CpGs (n = 1,961), methylation levels of which are significantly associated with the overall survival of patients in the test cohort and which were used for MethScore computation.

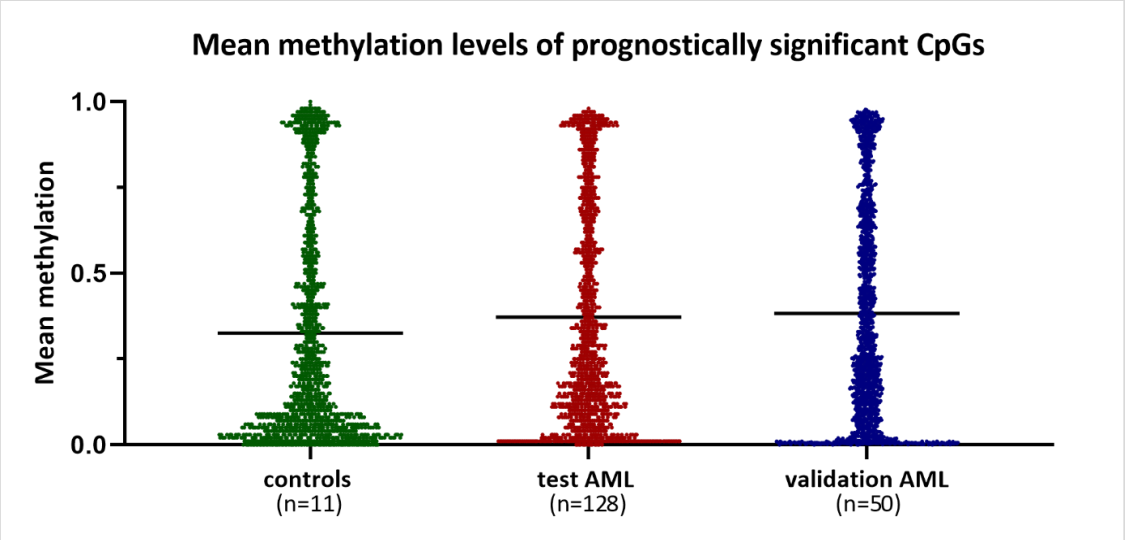


Figure 15 Mean methylation levels of CpGs (n = 1,961), methylation of which is significantly associated with the overall survival of patients in the test cohort, counted for the group of healthy donors' samples (controls), test cohort, and validation cohort of AML patients. The lines represent the mean of all values in each cohort.

In 864 CpGs from this selection, lower methylation was prognostically favorable. In the rest of the CpGs (n = 1,097) lower methylation indicated poorer survival. A total of 141 genes were annotated to these CpGs and these genes were mainly associated with DNA binding, regulation of RNA metabolism and transcription, and embryonic development. In **Table 21**, genes associated with CpGs most significant for patients' OS are shown.

Table 21 Genes annotated to most significant CpGs.

Top 10 significant CpGs from Cox univariate analysis	
Gene	**p-value**
HOTTIP	3.90E-05
RBL2	5.60E-05
EZH2 promoter	6.10E-05
AC012531.2	8.40E-05
HOXB3	1.00E-04
LTB	0.00012
HOXB7	0.00013
TNF	0.00013
HOTTIP	0.00013
EZH2 promoter	0.00013

HOTTIP - lncRNA associated with *HOXA* cluster, *AC012531.2* - lncRNA associated with *HOXC* cluster

Using the methylation levels and Cox regression coefficients of these 1,961 CpGs, we computed the MethScore (exact calculation is described in chapter 3.3.5) for each patient in the test cohort. MethScore was ranging from 72 to 1,565 with median 808 and average 814. To get an overview of acquired MethScore values, we computed the z-score and compared it with the overall average methylation and number of mutations of each patient, see **Figure 16**. It seems that higher MethScore slightly correlates with lower average methylation and higher number of mutations. We also estimated the MethScore of healthy donors' samples where the maximum value was 1,071 and minimum 462, median was 765, and mean 758.

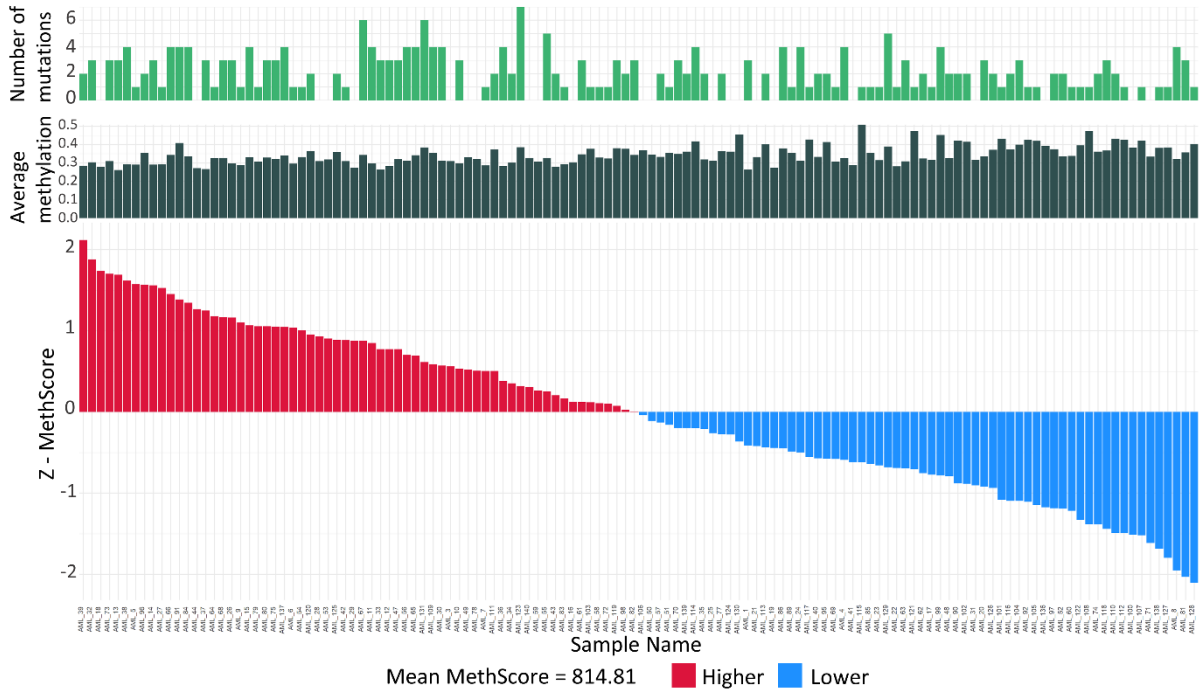


Figure 16 Z-score values computed from MethScore together with the average methylation and number of mutations for each patient from the test cohort (n = 128).

We subsequently divided the patients according to the median MethScore and compared the OS and EFS of the two groups. Patients with lower MethScore (n = 64) had markedly longer survival than patients with higher MethScore (n = 64, logrank test for OS: p < 2e-16; for EFS: p < 2e-12), see **Figure 17**.

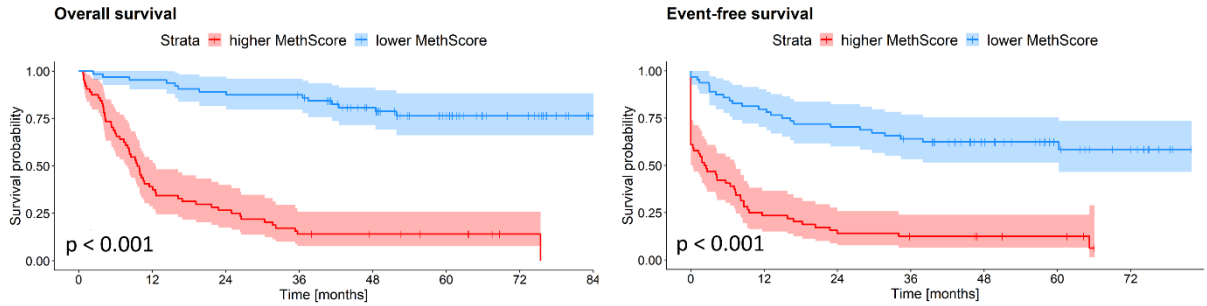


Figure 17 Kaplan-Meier curves with p-values of two-sided logrank test comparing both OS and EFS of patients with higher and lower MethScore in the test cohort (n = 128).

Next, we computed the MethScore for samples from the validation cohort to confirm its predictive ability for patients' survival. We used the same set of 1,961 CpGs and the same procedure (chapter 3.3.5). To see if there is any major difference in the methylation levels measured for the test and validation cohort, we compared the average methylation for each of the 1,961 CpGs. The values correlated well with $R^2 = 0.987$. The range of MethScore

values for the validation cohort was 244 – 1,413 with median 669, and mean 705. We computed the z-score and plotted it together with average methylation and number of mutations for each patient, the same way we did for the test cohort, see **Figure 18**. Patients were then divided according to the median MethScore and Kaplan-Meier curves were plotted (**Figure 19**). There was again a significant difference in both OS and EFS between patients with lower (n = 25) and higher (n = 25) MethScore.

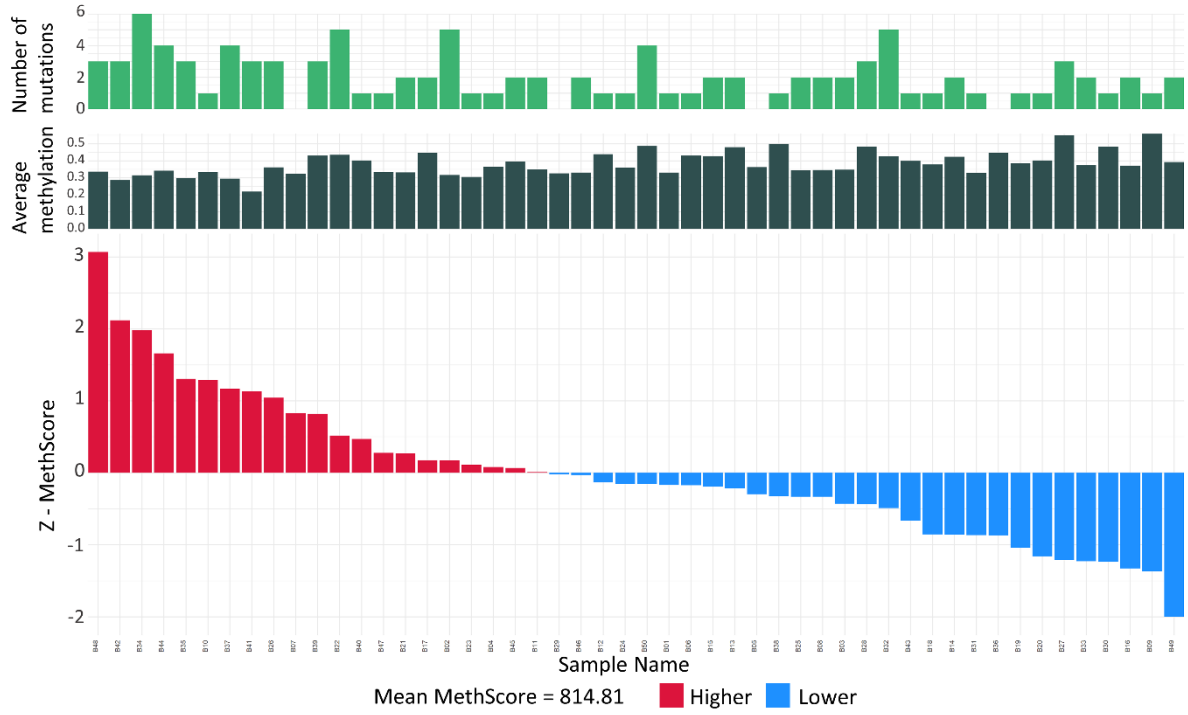


Figure 18 Z-score values computed from MethScore together with the average methylation and number of mutations for each patient from the validation cohort (n = 50).

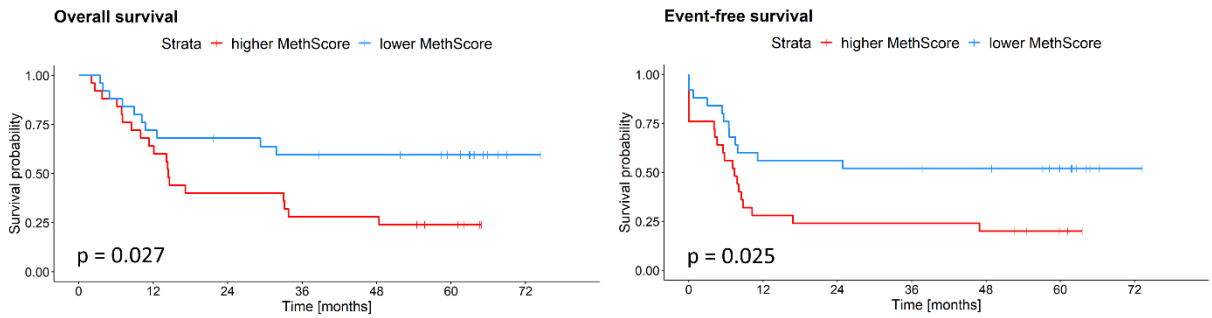


Figure 19 Kaplan-Meier curves with p-values of two-sided logrank test comparing OS and EFS of patients with higher and lower MethScore in the validation cohort (n = 50).

We further examined the prognostic relevance of MethScore in multivariate analyses. The results for both cohorts are summarized in **Tables 22** and **23** for OS and EFS, respectively. MethScore together with transplantation in the first remission and age were the most significant predictors of OS. *FLT3*-ITD mutation and number of leukocytes were

significant for OS only in the test cohort. On the contrary, the presence of *NPM1* mutation and cytogenetic classification were predictive for OS only for the validation cohorts' samples. MethScore together with the transplantation remained the most significant covariates also for EFS prediction. In the validation cohort, cytogenetic classification remained also highly significant and, in addition, mutations in *NPM1* came out as predictive. In the test cohort, *FLT3*-ITD together with mutations in *DNMT3A* and *NPM1* gained significance for EFS prediction. The hazard ratio of *DNMT3A* mutation would suggest a better outcome for patients with *DNMT3A*^{mut} which contradicts the current knowledge that classifies mutations in *DNMT3A* as an adverse prognostic factor (see chapter 1.2.2.1). However, it was shown that the outcome of *DNMT3A*^{mut} patients is strongly dependent on other co-mutations, especially *NPM1*^{mut} and *FLT3*-ITD (Gale *et al.*, 2015). In the test cohort, 67 % of patients with EFS shorter than 2 years (n = 74) had *NPM1*^{mut} and 42 % had *FLT3*-ITD, compared to patients with longer EFS (n = 54) where 60 % had *NPM1*^{mut} but only 27 % possessed *FLT3*-ITD. In **Figures 20** and **21**, full mutational backgrounds of investigated patients from both cohorts are shown.

Table 22 Summary of results from Cox multivariate regression analysis for overall survival

Overall Survival	Test cohort, n=128		Validation cohort, n=50	
Covariates	HR (95 % CI)	p-value	HR (95 % CI)	p-value
Age	1.05 (1.01-1.09)	**0.006**	1.05 (1.001-1.09)	**0.045**
Leukocyte count	1.01 (1.00-1.02)	**0.027**	0.99 (0.982-1.00)	0.105
Cytogenetics[1]	1.08 (0.45-2.60)	0.856	5.70 (1.307-24.87)	**0.021**
Transplantation in 1st CR	0.30 (0.13-0.69)	**0.005**	0.16 (0.046-0.58)	**0.005**
FLT3-ITD	2.34 (1.14-4.82)	**0.021**	2.01 (0.373-10.84)	0.417
DNMT3A mutation	0.53 (0.23-1.21)	0.134	3.36 (0.774-14.62)	0.106
IDH1/2 mutation	2.05 (0.97-4.32)	0.06	1.68 (0.440-6.41)	0.448
TET2 mutation	1.97 (0.72-5.37)	0.185	0.20 (0.020-2.06)	0.177
ASXL1 mutation	0.36 (0.11-1.22)	0.101	0.41 (0.085-1.97)	0.266
TP53 mutation	1.53 (0.52-4.54)	0.44	0.96 (0.122-7.59)	0.970
NPM1 mutation	0.46 (0.21-1.80)	0.053	0.21 (0.056-0.80)	**0.022**
CEBPA mutation	0.60 (0.20-1.80)	0.367	3.00 (0.326-27.54)	0.332
RUNX1 mutation	0.77 (0.23-2.61)	0.68	0.37 (0.029-4.68)	0.442
MethScore	1.01 (1.00-1.01)	**<0.001**	1.00 (1.002-1.01)	**0.002**

HR - hazard ratio, CI - confidence interval, CR - complete remission
[1] revised classification by Grimwade *et al.* (2010)

Table 23 Summary of result from Cox multivariate regression analysis for event-free survival.

Event-free Survival	Test cohort, n=128		Validation cohort, n=50	
Covariates	HR (95 % CI)	p-value	HR (95 % CI)	p-value
Age	1.01 (0.98-1.03)	0.589	1.02 (0.979-1.06)	0.390
Leukocyte count	1.00 (0.99-1.01)	0.832	0.99 (0.985-1.00)	0.267
Cytogenetics[1]	1.63 (0.77-3.46)	0.204	5.49 (1.519-19.85)	**0.009**
Transplantation in 1st CR	0.21 (0.10-0.42)	**<0.001**	0.07 (0.021-0.26)	**<0.001**
FLT3-ITD	2.78 (1.44-5.34)	**0.002**	1.40 (0.322-6.08)	0.654
DNMT3A mutation	0.34 (0.17-0.70)	**0.003**	2.45 (0.682-8.78)	0.170
IDH1/2 mutation	1.55 (0.82-2.93)	0.176	2.71 (0.697-10.57)	0.150
TET2 mutation	0.95 (0.39-2.31)	0.912	0.30 (0.046-1.94)	0.207
ASXL1 mutation	0.93 (0.36-2.41)	0.876	0.35 (0.078-1.55)	0.165
TP53 mutation	1.47 (0.52-4.16)	0.466	4.13 (0.53-32.23)	0.176
NPM1 mutation	0.49 (0.25-0.96)	**0.038**	0.27 (0.078-0.92)	**0.036**
CEBPA mutation	1.09 (0.43-2.77)	0.851	3.28 (0.375-28.69)	0.283
RUNX1 mutation	0.65 (0.25-1.72)	0.384	1.16 (0.109-12.33)	0.903
MethScore	1.00 (1.00-1.01)	**<0.001**	1.00 (1.001-1.01)	**0.003**

HR - hazard ratio, CI - confidence interval, CR - complete remission
[1] revised classification by Grimwade *et al.* (2010)

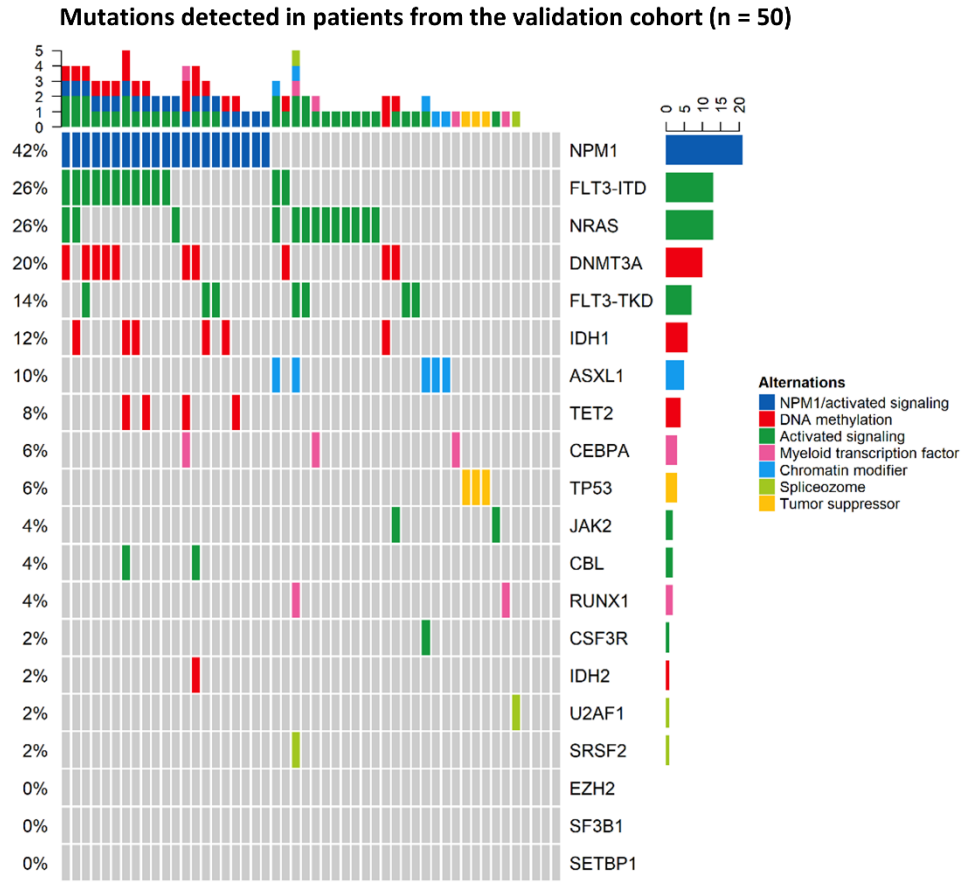


Figure 20 The mutational background of patients from the validation cohort investigated by NGS in collaborative study by Folta *et al.* (2019).

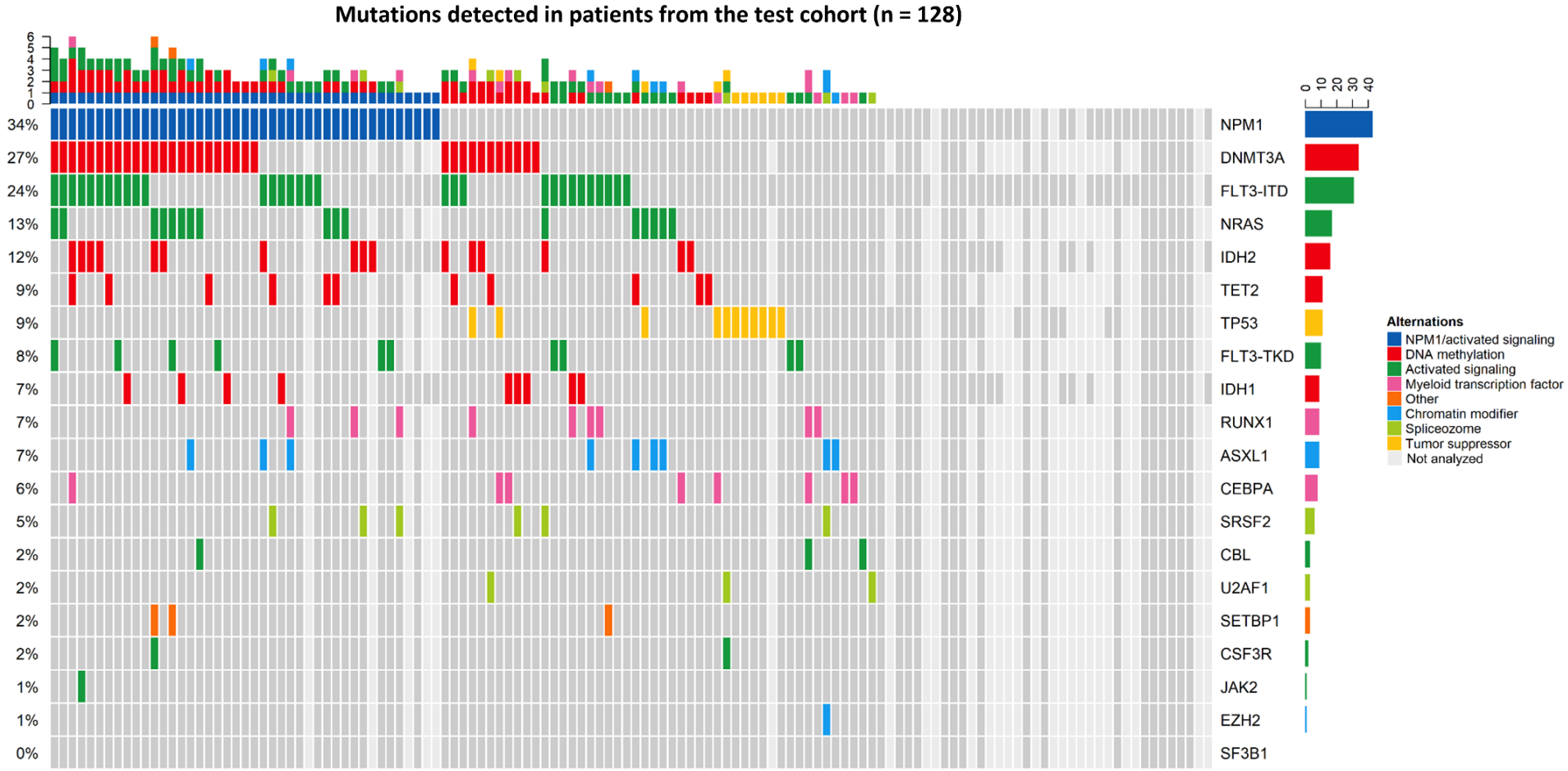


Figure 21 The mutational background of patients from the test cohort, NGS data were acquired in collaborative study by Folta *et al.* (2019), some of the patients were investigated in the Institute of Hematology and Blood Transfusion for *FLT3*-ITD and mutations in *NPM1*, *CEBPA*, *DNMT3A, IDH1/2* and *TP53* genes.

Additionally, in **Table 24** we provide further comparison of patients with lower and higher MethScore for both cohorts. As expected, the most significant differences were in the survival time and number of deaths. Furthermore, in the test cohort, the two groups differed in the representation of AML subtypes. There was significantly more M2 type AML in the low MethScore group and more M4 AML in the high MethScore group. The majority of patients with favorable cytogenetics were in the lower MethScore group and, opposingly, most of the patients with adverse cytogenetic changes were those with higher MethScore. The groups also differed in the number of transplanted patients and patients that reached complete remission that were more represented in the lower MethScore group. Nearly all patients with mutated *TP53* and majority of patients with mutations in *TET2* were those with higher MethScore. Patients with lower MethScore were also significantly younger, had lower levels of leukocytes at diagnosis, and also fewer gene mutations when compared with higher MethScore group. In the validation cohort, the groups varied less. There was a significant difference in the representation of male and female patients. Furthermore, nearly all patients with *FLT3*-ITD were in the higher MethScore group and patients from this group had also significantly more mutations than those with lower MethScore.

Table 24 Comparison of patients with lower (< median) and higher (> median) MethScore.

		Test cohort (n=128)			Validation cohort (n=50)		
		Lower MethScore (n = 64)	**Higher MethScore (n = 64)**	**p-value**	**Lower MethScore (n = 25)**	**Higher MethScore (n = 25)**	**p-value**
FAB classification	M1	18	13	**0.028**	6	1	0.451
	M2	19	7		4	6	
	M3	0	0		0	0	
	M4	10	20		4	6	
	M5	6	11		1	2	
	M6	4	2		2	0	
	M7	0	1		0	0	
	AML RAEB	7	10		8	9	
Cytogenetics (Grimwade 2010)	Favorable	8	1	**0.001**	2	1	1
	Intermediate	47	40		19	19	
	Adverse	8	22		4	5	
Sex	Male / Female	34 / 30	34 / 30	1	16 / 9	8 / 17	**0.046**
Transplantation in 1st CR	Yes / No	36 / 28	21 / 43	**0.012**	12 / 13	6 / 19	0.140
Relapse		20 / 44	22 / 42	0.851	10 / 15	13 / 12	0.571
CR after 1st induction		46 / 18	24 / 40	**0.0003**	17 / 8	17 / 8	1
***FLT3*-ITD**	Positive / Negative / NA	13 / 50 / 1	19 / 45 / 0	0.307	2 / 23	11 / 14	**0.008**
***DNMT3A* mutation**		17 / 41 / 6	23 / 32 / 9	0.175	3 / 22	7 / 18	0.289
***IDH1/2* mutation**		12 / 46 / 6	14 / 41 / 9	0.656	4 / 21	3 / 22	1
***TET2* mutation**		2 / 51 / 11	9 / 40 / 15	**0.024**	1 / 24	3 / 22	0.609
***ASXL1* mutation**		4 / 49 / 11	5 / 44 / 15	0.735	1 / 24	4 / 21	0.349
***NRAS* mutation**		6 / 47 / 11	11 / 38 / 15	0.184	7 / 18	6 / 19	1
***TP53* mutation**		1 / 53 / 10	11 / 45 / 8	**0.004**	2 / 23	1 / 24	1
***NPM1* mutation**		19 / 43 / 2	24 / 37 / 3	0.348	9 / 16	12 / 13	0.567
***CEBPA* mutation**		4 / 59 / 1	4 / 55 / 5	1	1 / 24	2 / 23	1
***RUNX1* mutation**		4 / 49 / 11	5 / 44 / 15	0.735	1 / 24	1 / 24	1

		Test cohort (n=128)			Validation cohort (n=50)		
		Lower MethScore (n = 64)	**Higher MethScore (n = 64)**	**p-value**	**Lower MethScore (n = 25)**	**Higher MethScore (n = 25)**	**p-value**
Number of mutations	Average ± SD / Median	1.6 ± 1.3 / 1	2.3 ± 1.7 / 2.5	**0.021**	1.6 ± 1.0 / 1	2.5 ± 1.6 / 2	**0.034**
Age		46.4 ± 13.7 / 49.1	54.2 ± 10.4 / 57.85	**0.001**	57.7 ± 13.9 / 62	53.4 ± 13.7 / 57	0.180
Leukocytes		59 ± 38.6 / 53	74.1 ± 39.8 / 78	**0.032**	46.3 ± 61.0 / 19.9	51.9 ± 60.0 / 25.9	0.318
Blasts PB		30.7 ± 31.18 / 17.4	34.1 ± 35.4 / 19	0.770	33.0 ± 32.3 / 18	37.3 ± 30.1 / 27	0.869
Blasts BM		52.7 ± 26.2 / 53.6	54.9 ± 28.8 / 52.8	0.601	49.7 ± 23.9 / 45	53.6 ± 22.1 / 54.9	0.542
OS [months]		52.1 ± 20.1 / 51.9	17.8 ± 19.5 / 9.7	**< 0.0001**	40.4 ± 26.2 / 51.8	26.3 ± 22.6 / 14.4	0.061
EFS [months]		40.7 ± 25.2 / 45.0	11.2 ± 18.4 / 2.3	**< 0.0001**	34.3 ± 28.2 / 37.7	17.5 ± 22.8 / 7.4	**0.029**
Death	Event / Censored	14 / 50	56 / 8	**< 0.0001**	10 / 15	19 / 6	**0.021**
Death/Relapse		25 / 39	57 / 7	**< 0.0001**	12 / 13	20 / 5	**0.038**

FAB - French-American-British classification, CR - complete remission, NA - not analyzed, SD - standard deviation, PB - peripheral blood, BM - bone marrow

4.3.2.1 MethScore in patients from intermediate risk group

We wanted to evaluate the prognostic capability of MethScore for AML patients from intermediate risk group because especially they would benefit from a better prognostic stratification. To get statistically more significant results, we combined patients from both cohorts and selected those with intermediate cytogenetic risk (according to Grimwade *et al.,* 2010). We further divided the patients into four groups: patients with intermediate cytogenetics (IR, n = 112; 74 from the test cohort, 38 from the validation cohort); patients with intermediate cytogenetics and *FLT3*-ITD (IR^{FLT3+}, n = 37; 25 from the test cohort, 12 from the validation cohort); patients with intermediate cytogenetics and wild-type *NPM1* (IR^{NPM1-}, n = 60; 40 from the test cohort, 20 from the validation cohort); and patients with intermediate cytogenetics, *FLT3*-ITD and $NPM1^{mut}$ ($IR^{FLT3+NPM1+}$, n = 23; 13 from the test cohort, 10 from the validation cohort). In each group, patients with lower and higher MethScore, as estimated by median values in the previous analysis, were compared.

It is evident from the Kaplan-Meier curves shown in **Figures 22** and **23** that MethScore reliably stratified patients with better and worse survival. We further validated the prognostic ability of MethScore in multivariate analyses. Because of the small number of patients, it was not possible to perform the multivariate analysis for the $IR^{FLT3+NPM1+}$ group. Results for the other groups of patients are summarized in **Tables 25** and **26**. MethScore was the most significant variable for both OS and EFS together with the transplantation in the first remission. In the group of IR patients, other significant covariates were number of leukocytes at diagnosis and mutations in *NPM1* which were significant for both OS and EFS. *FLT3*-ITD came out as significant only for EFS. In the IR^{NPM1-} group, age, leukocytes, and mutations in *IDH1/2* were predictive of patients' EFS. Mutations in *TET2* and *ASXL1* together with age were among significant covariates for EFS in IR^{FLT3+} group. The hazard ratio incorrectly suggests that $ASXL1^{mut}$ is a favorable marker for patients' outcome, which it is not (see chapter 1.2.2.1). However, only three patients in the IR^{FLT3+} had *ASXL1* mutations, so the prognostic effect could not be appropriately evaluated. Also, the hazard ratio of age, which was a significant variable for EFS in IR^{FLT3+} and IR^{NPM1-}, suggest an opposite effect than expected. However, this may be caused by the fact that IR patients were mostly younger (47 ± 14 years in IR^{FLT3+}, 52 ± 13 in IR^{NPM1-}) and thus the consideration of age may not be entirely accurate.

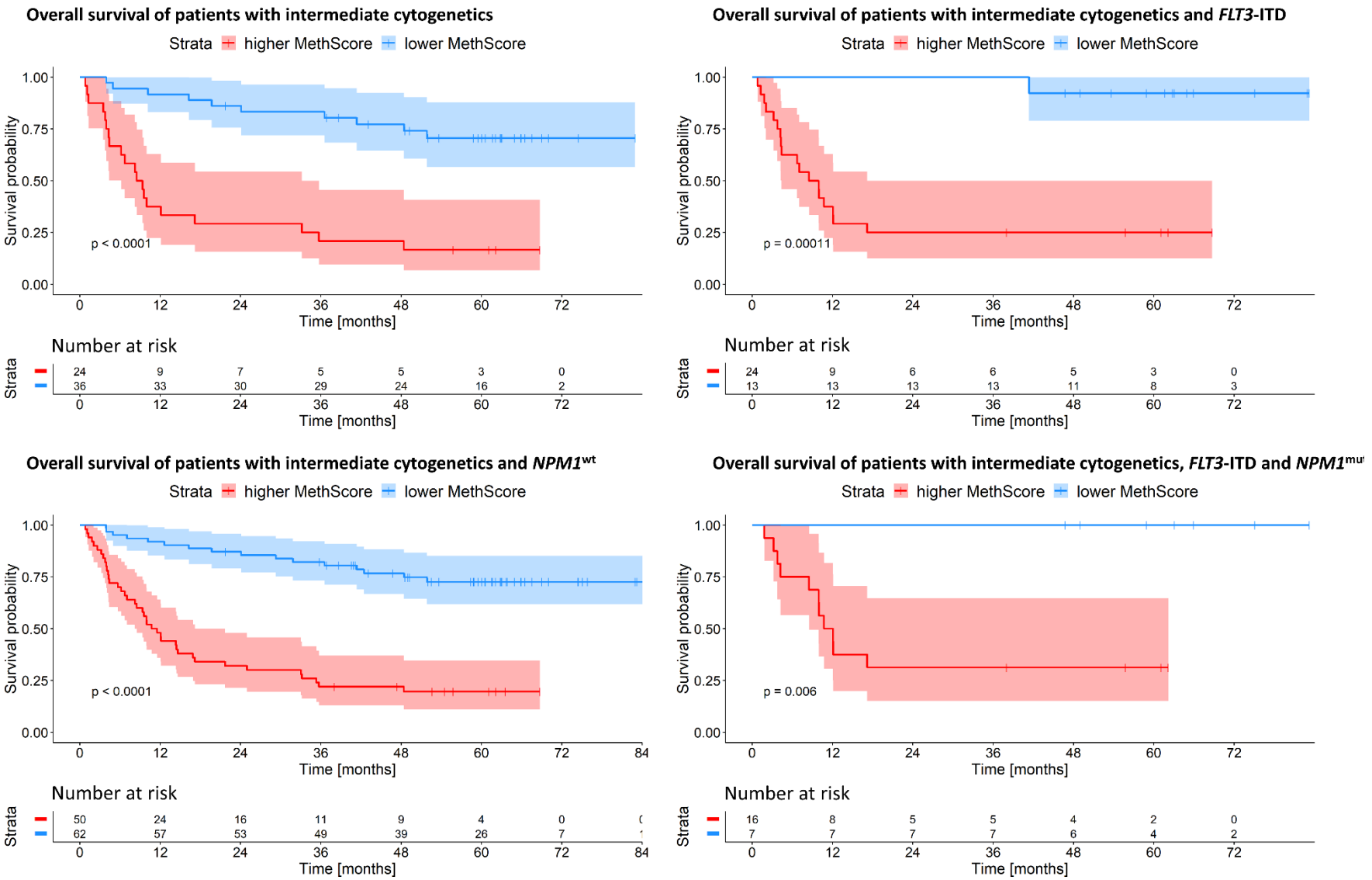


Figure 22 Kaplan-Meier curves for overall survival of patients with intermediate risk and lower or higher MethScore

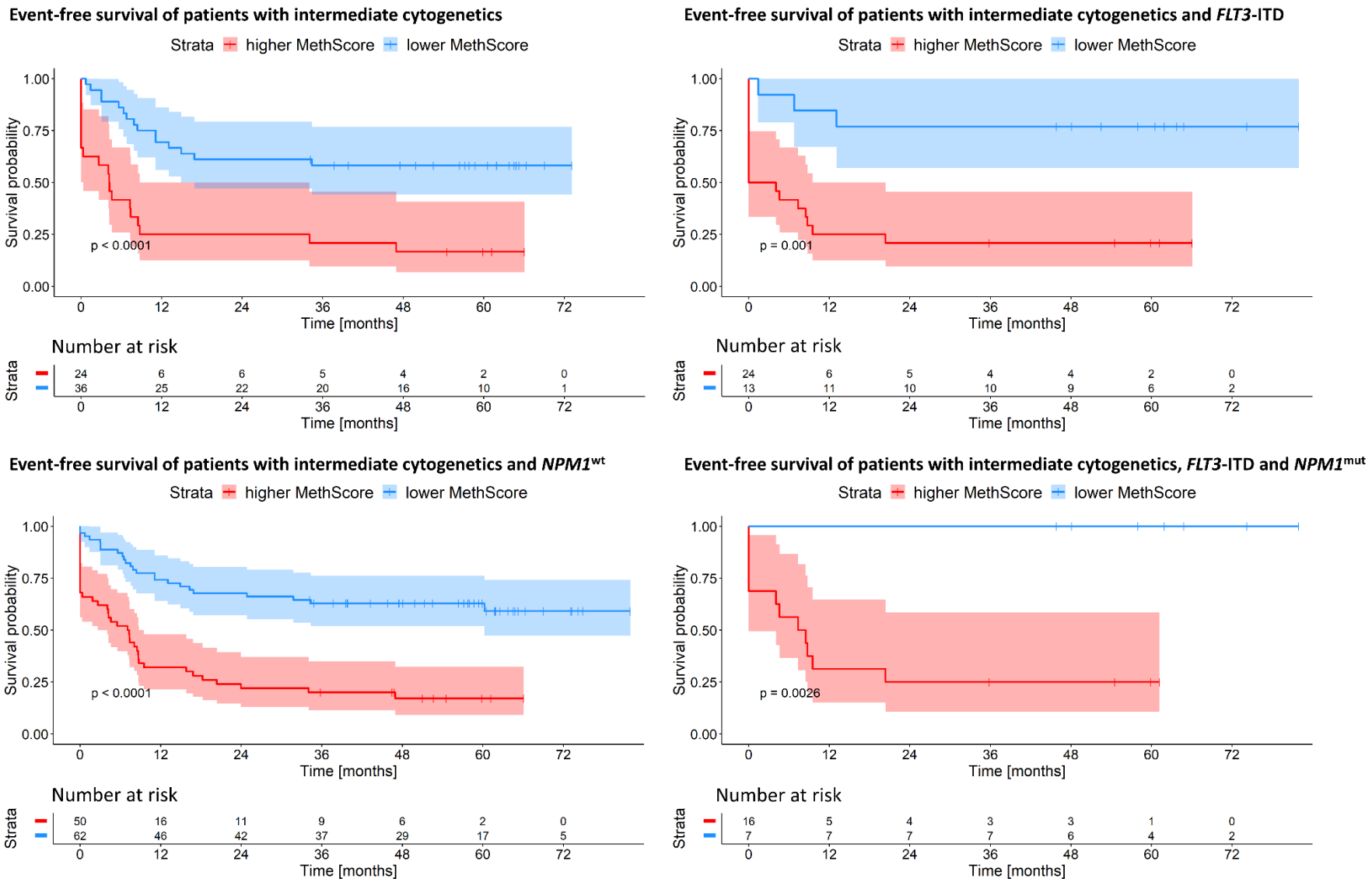


Figure 23 Kaplan-Meier curves for event-free survival of patients with intermediate risk and lower or higher MethScore

Table 25 Results from multivariate Cox regression analysis of overall survival of patients from intermediate risk groups

Overall Survival	Intermediate cytogenetics (n=112)		Intermediate cytogenetics, *FLT3*-ITD (n=37)		Intermediate cytogenetics, *NPM1*^{wt} (n=50)	
Covariates	HR (95 % CI)	p-value	HR (95 % CI)	p-value	HR (95 % CI)	p-value
Age	1.012 (0.979-1.05)	0.48	0.951 (0.895-1.01)	0.105	0.952 (0.899-1.01)	0.086
Leukocyte count	0.991 (0.984-1.00)	**0.012**	0.989 (0.976-1.00)	0.118	0.991 (0.982-1.00)	0.065
Transplantation in 1st CR	0.097 (0.034-0.28)	**<0.001**	0.046 (0.008-0.27)	**<0.001**	0.069 (0.014-0.35)	**0.001**
FLT3-ITD	2.167 (0.973-4.82)	0.058	-	-	0.686 (0.159-2.96)	0.614
DNMT3A mutation	0.974 (0.440-2.16)	0.949	0.516 (0.099-2.70)	0.433	0.913 (0.247-3.37)	0.892
IDH1/2 mutation	0.813 (0.374-1.77)	0.602	2.044 (0.448-9.34)	0.356	1.336 (0.459-3.89)	0.595
TET2 mutation	0.617 (0.188-2.03)	0.426	0.265 (0.040-1.75)	0.168	0.365 (0.070-1.92)	0.234
ASXL1 mutation	0.369 (0.087-1.56)	0.174	0.043 (0.001-1.48)	0.081	0.243 (0.005-12.97)	0.485
TP53 mutation	0.571 (0.04-8.09)	0.678	-	-	0.903 (0.043-19.17)	0.948
NPM1 mutation	0.327 (0.153-0.70)	**0.004**	0.619 (0.178-2.15)	0.451	-	-
CEBPA mutation	1.611 (0.325-7.99)	0.559	-	-	0.894 (0.116-6.92)	0.914
RUNX1 mutation	0.349 (0.066-1.86)	0.217	0.548 (0.019-16.14)	0.727	0.357 (0.009-13.87)	0.581
MethScore	1.005 (1.003-1.01)	**< 0.001**	1.008 (1.004-1.01)	**<0.001**	1.007 (1.004-1.01)	**<0.001**

HR - hazard ratio, CI - confidence interval, CR - complete remission
Values are missing for genes that were mutated/wild-type for all patients in the tested group.

Table 26 Results from multivariate Cox regression analysis of event-free survival of patients from intermediate risk groups

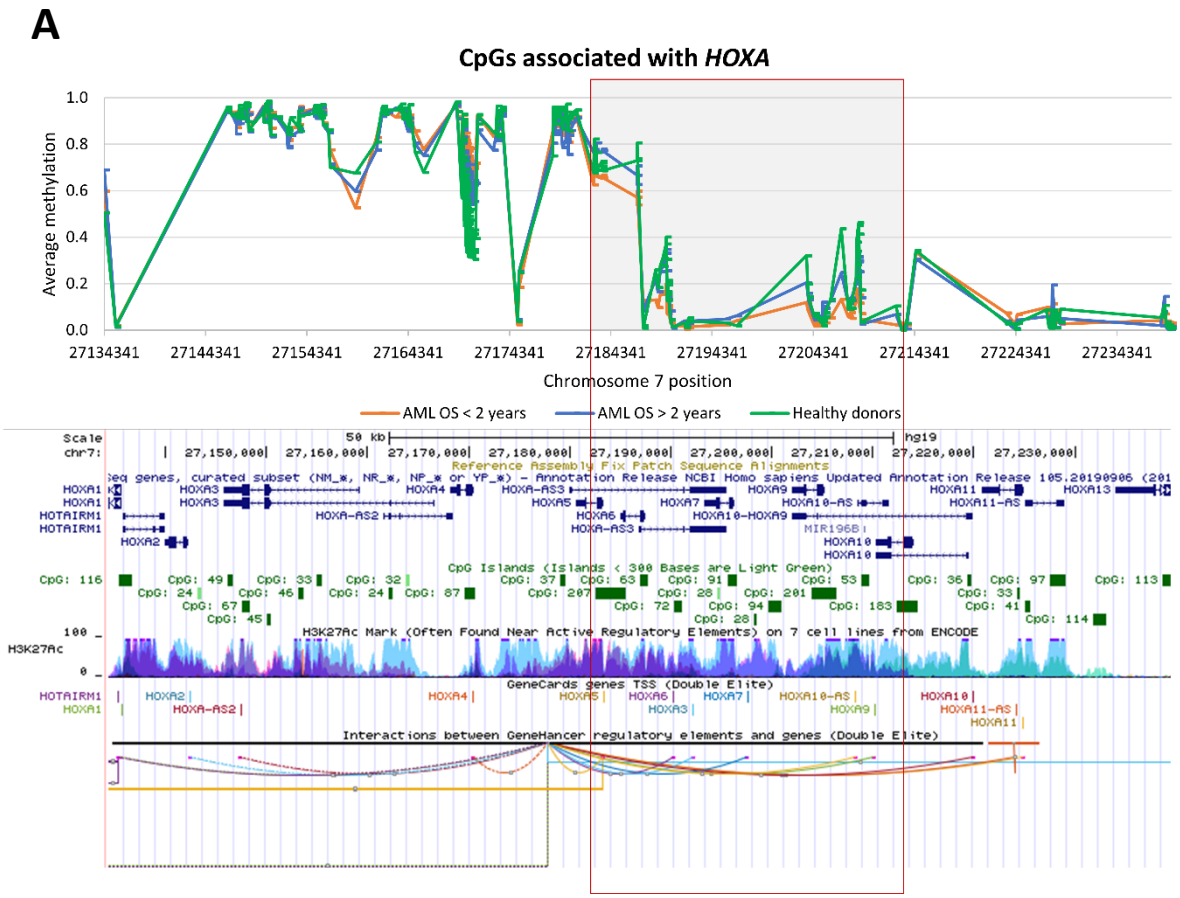
Event-free Survival	Intermediate cytogenetics (n=112)		Intermediate cytogenetics, *FLT3*-ITD (n=37)		Intermediate cytogenetics, *NPM1*^{wt} (n=50)	
Covariates	HR (95 % CI)	p-value	HR (95 % CI)	p-value	HR (95 % CI)	p-value
Age	0.99 (0.964-1.02)	0.422	0.927 (0.879-0.98)	**0.005**	0.929 (0.887-0.97)	**0.001**
Leukocyte count	0.99 (0.986-1.00)	**0.018**	0.989 (0.977-1.00)	0.102	0.990 (0.982-1.00)	**0.027**
Transplantation in 1st CR	0.11 (0.046-0.25)	**<0.001**	0.056 (0.012-0.27)	**<0.001**	0.046 (0.012-0.18)	**<0.001**
FLT3-ITD	3.12 (1.538-6.33)	**0.002**	-	-	1.600 (0.529-4.84)	0.406
DNMT3A mutation	0.63 (0.303-1.32)	0.223	0.226 (0.046-1.11)	0.067	0.593 (0.190-1.85)	0.369
IDH1/2 mutation	1.08 (0.555-2.11)	0.818	0.927 (0.304-2.83)	0.894	2.764 (1.068-7.16)	**0.036**
TET2 mutation	0.51 (0.185-1.41)	0.195	0.113 (0.020-0.64)	**0.013**	0.370 (0.090-1.52)	0.168
ASXL1 mutation	0.56 (0.200-1.55)	0.263	0.097 (0.012-0.79)	**0.029**	1.478 (0.134-16.32)	0.75
TP53 mutation	2.22 (0.187-26.43)	0.527	-	-	12.783 (0.655-249.63)	0.093
NPM1 mutation	0.35 (0.175-0.71)	**0.003**	0.748 (0.243-2.3)	0.613	-	-
CEBPA mutation	2.79 (0.730-10.66)	0.134	-	-	0.979 (0.156-6.15)	0.982
RUNX1 mutation	0.57 (0.175-1.88)	0.358	0.407 (0.071-2.31)	0.310	0.281 (0.033-2.36)	0.242
MethScore	1.00 (1.003-1.01)	**<0.001**	1.007 (1.003-1.01)	**<0.001**	1.005 (1.003-1.01)	**<0.001**

HR - hazard ratio, CI - confidence interval, CR - complete remission
Values are missing for genes that were mutated/wild-type for all patients in the tested group.

4.3.3 *HOX* genes

In the set of 1,961 CpGs that were used for the MethScore computation, 637 CpGs (32.5 %) were associated with *HOX* genes. Most CpGs belonged to the *HOXA* gene cluster (n = 294) and here, CpGs for which lower methylation values indicated better patients' outcome prevailed (76 %). *HOXB* associated CpGs were also highly represented (n = 149) and there was nearly an equal number of CpGs with prognostically positive lower (51 %) and higher (49 %) methylation values. The rest of the significant CpGs were assigned to *HOXC* (n = 33) and *HOXD* (n = 161) gene clusters and majority of these CpGs (73 % and 80 %, respectively) were those for which hypermethylation was favorable for patients' outcome.

To better understand the observed DNA methylation changes in case of *HOXA* and *HOXB* genes, we plotted the average methylation values of healthy donors and AML samples divided according to their survival, see **Figure 24**. We revealed a distinct regulatory region in both *HOX* clusters where patients with shorter survival were clearly hypomethylated.



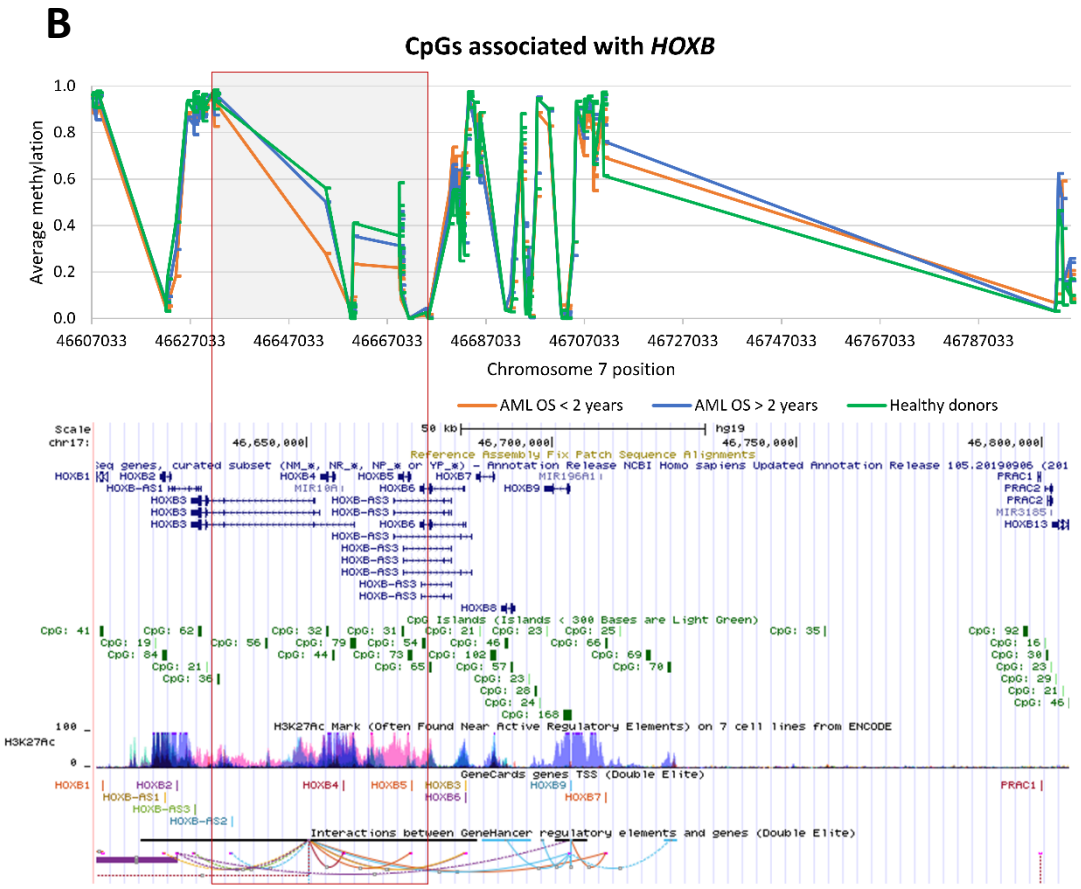


Figure 24 Average methylation levels of CpGs within *HOXA/B* clusters that were found as significant for patients' OS. Values for healthy donors (n = 11), AML patients with OS < 2 years (n = 54), and AML patients with OS > 2 years (n = 74) are shown. The lower half of each image was taken from UCSC Genome Browser, assembly GRCh38/hg19 (Kent *et al.*, 2002) **A** *HOXA* gene cluster **B** *HOXB* gene cluster. Regions with hypomethylation in patients with shorter survival are highlighted.

5 DISCUSSION

5.1 DNA methylation and hydroxymethylation changes in AML patients with mutations in *DNMT3A*, *IDH1/2* or their combinations

In this project, we wanted to evaluate the DNA methylation, hydroxymethylation and expression landscape of patients with distinct mutational background. We focused on patients with mutations in *DNMT3A* and *IDH1/2* or their combinations, because these genetic aberrations have an opposing effect on DNA methylation (Russler-Germain *et al.*, 2014; Fong *et al.*, 2014) and especially the epigenetic landscape of patients with both genes concurrently mutated may be interesting.

We investigated the overall methylation and hydroxymethylation profiles via EPIC arrays (Illumina) that target nearly 850 thousand CpGs throughout the genome. The results showed a profound hypomethylation of DNA in *DNMT3A*^{mut} patients and a prevailing DNA hypermethylation in *IDH1/2*^{mut} patients, which is in accordance with previous studies (Figueroa *et al.*, 2010; Hájková *et al.*, 2012). It is noticeable, that there was a significant difference in number of DMPs between *IDH1*^{mut} and *IDH2*^{mut} patients. This somewhat disagrees with a study that described highly similar DNA methylation profiles for *IDH1* and *IDH2* mutated patients (Figueroa *et al.*, 2010). However, in a more recent paper, the theory that the biological background of those two mutations is distinct was published in case of gliomas (Wang *et al.*, 2016). Patients with mutations in both genes (*DNMT3A&IDH1/2*^{mut}) had markedly fewer differentially methylated sites and the numbers of hyper- and hypomethylated DMPs were more balanced than in patients with only one gene mutated. *DNMT3A&IDH1/2*^{mut} patients thus exhibited a mixed DNA methylation profile that reflected the antagonistic influence of the two mutated genes. The same conclusions were presented in a recent study, where authors further reported that most CpGs hypermethylated in *IDH2*^{mut} patients and hypomethylated in *DNMT3A*^{mut} patients were unchanged in *DNMT3A&IDH1*^{mut} patients and samples with both genes mutated had a specific set of hypo- and hypermethylated CpGs (Glass *et al.*, 2017).

Regarding the DNA hydroxymethylation, we detected a significantly lower number of hydroxymethylated sites in all samples' groups when compared with aberrantly methylated CpGs. The lowest number of 5-hmCs was detected among *DNMT3A*^{mut} patients. This may be explained by the substantial DNA hypomethylation in those samples, causing a relative absence of sites where hydroxymethylation can occur via demethylation processes

(Ko *et al.*, 2015). Surprisingly, the highest level of hydroxymethylation was observed in *IDH2*^{mut} patients. This is a rather unexpected observation when we consider that it is the product of mutated IDH1/2 enzyme that inhibits the TET2-mediated DNA hydroxymethylation (Grimwade *et al.*, 2016). However, *IDH2*^{mut} patients had the highest number of hypermethylated CpGs that could be oxidized via TET2 so it is possible that despite the decreased activity of the enzyme, there was still the biggest opportunity for hydroxymethylation to occur. Nevertheless, we are not able to drive a clear assumption regarding the DNA hydroxymethylation patterns, due to the generally low numbers of 5-hmCs detected in our blood samples. This also caused difficulties during the validation of array data, where we were unable to properly measure the 5-hmC levels with pyrosequencing. To confirm our observations regarding the low levels of DNA hydroxymethylation, we performed an additional experiment where we measured the total amount of methylated and hydroxymethylated cytosines in blood samples via mass spectrometry. Our results agreed with a study comparing DNA hydroxymethylation in samples from brain and blood tissue, where authors acquired the same low levels of 5-hmC and 5-mC for a blood sample (0.04 % and ~ 4 %, respectively) as we did (Stewart *et al.*, 2015).

Next, we focused on the influence of other mutations on the overall DNA hydroxy-/methylation patterns. We noticed an effect of *NPM1* mutations in the clustering analysis of DNA methylation data which is in line with previously described observations of *NPM1*^{mut} influencing the hypermethylation phenotype in AML patients (Figueroa *et al.*, 2010). We did not detect any impact of *TET2* mutations on DNA methylation. The reason why samples with *DNMT3A*^{mut} and *TET2*^{mut} did not resemble *DNMT3A&IDH1/2*^{mut} samples may lie in the different degree of TET2 inhibition via 2-hydroxyglutarate compared to a direct mutation of the enzyme. Nevertheless, we observed an apparent influence of *TET2*^{mut} on the DNA hydroxymethylation pattern.

Our results further showed that the overall expression profiles of investigated samples do not copy the layout of either DNA methylation or hydroxymethylation patterns. Therefore, we could assume that not all DNA hydroxy-/methylation alterations are translated into changes in gene expression. Similar findings were published by Spencer *et al.* (2017), where the authors did not find any clear trends when comparing the gene expression and DNA methylation in associated gene promoters, gene bodies or other regulatory elements in AML patients with *DNMT3A*^{mut}. This is further supported by a recent study showing the great complexity of the interplay between DNA methylation and gene expression in cancer

(Spainhour *et al.*, 2019). Nevertheless, we did observe a slight connection because the most differentially expressed genes were detected in *DNMT3A*^{mut} and *IDH2*^{mut} samples that also possessed the highest number of DMPs. On the contrary, least deregulated gene expression was observed among *DNMT3A&IDH2*^{mut} samples, methylation profiles of which were the ones most similar to healthy donors.

5.1.1 Prognostically significant DNA methylation changes

We investigated the prognostic significance of discovered aberrant DNA methylation in two regions. The first site was associated with tumor suppressor *CHFR*. The effect of differential methylation and expression of *CHFR* on patients' outcome was discussed in two recent papers (Gao *et al.*, 2016; Zhou *et al.*, 2018). In one publication, hypermethylation of the *CHFR* promoter region was described as a frequent event (observed in 24 % of samples) associated with an adverse outcome of AML patients (Gao *et al.*, 2016). We did not confirm this observation with our data as we detected higher methylation levels only in a small fraction of patients (4 %). Furthermore, we found out that the region studied by Gao *et al.* (2016) actually belonged to a promoter of another gene, *ZNF605*, located nearly 70 kbp upstream from the *CHFR* promoter. Thus, the results are not comparable with our data. In the second study, authors claimed that DNA methylation in *CHFR* promoter (a site adjacent to our investigated region) is a rare event and is not connected with downregulation of *CHFR* expression, which is a negative prognostic factor for AML patients (Zhou *et al.*, 2018). This is in accordance with our findings regarding the DNA hypermethylation; however, we also observed a strong connection with *CHFR* downregulation. The prognostic effect of the DNA hypermethylation and the corresponding decrease in gene expression could not be evaluated in our case due to the small number of affected patients.

The second investigated region belonged to an enhancer of *GZMB* gene encoding granzyme B, a protein indispensable for the immune response of natural killer (NK) cells and T-cells (Cao *et al.*, 2007). We discovered a negative prognostic effect of DNA hypermethylation in the area that was independent of blast percentage and *IDH1/2* mutations. However, we did not find a clear association between the increased methylation and expression of *GZMB*. One explanation may be that the enhancer located 40 kbp upstream from *GZMB* gene is associated with another gene. *GZMB* is located on chromosome 17 and tightly linked with *GZMH* (granzyme H) and *CTSG* (cathepsin G) genes located in the close proximity (Sedelies *et al.*, 2004). The GeneHancer, a database of human enhancers and their probable target sites, indicated a connection of our hypermethylated site with all three genes.

Nevertheless, a gene ontology tool GREAT linked our region only to granzyme B. It is also possible that *GZMB* transcription is affected, however, we are not able to detect it in the whole blood mRNA samples since active expression of *GZMB* occurs only in NK-cells and cytotoxic T-cells (Susanto *et al.*, 2012). The disruption of the immune response mechanism would explain the inferior outcome of *GZMB* hypermethylated patients.

5.2 DNA methylation validation methods

In order to choose the best method for DNA methylation validation, needed in our first project, we decided to compare four most common methods used for these purposes: analysis with methylation dependent or independent restriction enzymes (MSRE), pyrosequencing, methylation-specific high-resolution melting analysis (MS-HRM), and methylation-specific quantitative PCR (qMSP).

MSRE analysis proved to be the fastest and easiest method. However, it provided accurate measurements only for either highly methylated or unmethylated sites. MSRE cleavage was not suitable for proper assessment of intermediately methylated regions for which we obtained underestimated values. We found out that shorter cleavage time (one hour instead of two) did not improve this disability, however, it made the analysis even faster without spoiling the results for hyper- or unmethylated sites. One more disadvantage of this method is that it is suitable only for CpGs placed inside a specific restriction sequence and at least two restriction sites must be present in the studied DNA region (Itoi *et al.*, 2007). Thus, it is impossible to establish the methylation level of a single CpG.

The only method with base-pair resolution was pyrosequencing. This advantage is somehow counterweighted with the ability to study only short (100 - 150 bp) DNA regions (Delaney *et al.*, 2015). This limitation is caused by the methodology itself because with every dispensation cycle the volume of the reaction increases and the background signal rises (Tost & Gut, 2007). We confirmed this in our experiments with an observation of decreasing quality of the signal after 45th dispensation cycle in the resulting pyrograms which corresponded to ~ 100 bp sequence. Another benefit of pyrosequencing is the use of bisulfite conversion control that enables to check for complete BS conversion (Tost & Gut, 2007). We always included at least three BS controls in every pyrosequencing assay. Altogether, we confirmed the benefits of pyrosequencing published by Reed *et al.* (2010). According to this paper, pyrosequencing may replace the bisulfite sequencing (via cloning) which is still considered a golden standard for DNA methylation analysis (Reed *et al.*, 2010; Al Harrasi *et al.*, 2017).

MS-HRM analysis was easy to perform and accurate in DNA methylation assessment. The biggest advantage of MS-HRM over pyrosequencing is the low price of the method. It is also recommended for this approach to keep the amplicon small in order to reduce the complexity of resulting melting profiles (Hansen *et al.*, 2008). However, the shorter PCR product is used, the higher the sensitivity but lower the resolution of the method becomes. The methylation profiles of longer sequences are more distinguishable (Wojdacz & Dobrovic, 2007). In our experiments, we were also addressing the question whether to include a CpG site into the primers' sequence or not. The study group of T.K. Wojdacz repeatedly showed that CpGs present in the 5'-end area of primers compensate a PCR bias towards unmethylated DNA sequence by favoring the amplification of methylated sequence, resulting in an increased sensitivity of the method (Wojdacz & Hansen, 2006; Wojdacz *et al.*, 2008; Wojdacz *et al.*, 2009). In our experiments, we compared the results from MS-HRM acquired with primers designed either with or without CpGs in their sequence. The Wojdacz modification resulted in more accurate calibration curves for the unmethylated region for both AUC-based and peak height-based approaches. However, it had an exactly opposite effect on calibration curves for methylated region, where the accuracy dropped. Moreover, in the AUC-based assessment the Wojdacz modification resulted in a disproportionately high methylation value for the methylated region. Thus, it seems that introducing CpGs inside primers' sequences is not so straightforwardly beneficial and it apparently requires and additional thorough optimization of annealing temperature and PCR conditions.

The last evaluated method was qMSP. This method, similarly to MSRE analysis, seemed to be more suitable for highly methylated or unmethylated regions but not accurate enough for measurement of intermediate DNA methylation. The overall precision of the method was also low and the introduction and optimization was very demanding. We do not consider this method to be suitable for proper measurement of DNA methylation. Unfortunately, it is still a quite popular method among researchers used in many studies (Zhou *et al.*, 2016; Guo *et al.*, 2017; Zhou *et al.*, 2017; Li *et al.*, 2017).

In the overall comparison, pyrosequencing proved to be the best method, not demanding and precise with high resolution, with the only disadvantage being the high price of the instrument. MS-HRM provided comparably accurate results as pyrosequencing and is affordable and easily performed. MSRE analysis was the fastest and simplest approach but suitable only for highly methylated or unmethylated regions. Similarly, qMSP with its several

drawbacks may better serve only as a method for decision whether the region is or is not methylated rather than for an assessment of an exact DNA methylation level.

5.3 DNA methylation sequencing panel

5.3.1 Comparison of our sequencing data with source literature

We based our custom DNA methylation panel on twenty previously published studies that described specific DNA methylation changes with an impact on patient's outcome. With our data, we tried to validate the prognostic significance reported in sixteen of them. We were unable to reproduce and validate the results of four studies that developed a specific approach for the evaluation of DNA methylation changes. Figueroa *et al.* (2010) used a novel methylation classifier based on principal component analysis. However, we were not able to get the SuperPC tool (Bair *et al.*, 2006) to work to repeat their analysis. This approach was further developed by Wertheim *et al.* (2015) and Luskin *et al.* (2016). Using random forest method, the authors computed a value called M-score. We did not perform the calculation of M-score due to its methodological complexity. And finally, Li *et al.* (2016) evaluated the shift of DNA methylation in the genome rather than the actual methylation levels. Nevertheless, we tried to reproduce the results for the rest of the studies as faithfully and meaningfully as possible.

We were able to confirm the reported prognostic significance in case of only five publications. One possible explanation is the difference in methodologies used for the DNA methylation assessment. As discussed in the previous chapter, qMSP may not be a very accurate method. Nevertheless, it was used in six publications (when we include MethyLight as a qMSP method) and we did not validate the results from any of them. Apart from the methodical issues, a lack of a validation cohort of samples may be another reason for the validation failure. From the sixteen studies evaluated, only seven verified their results on an independent set of patients. From these seven, we confirmed the results of four publications (Božić *et al.*, 2015; Deneberg *et al.*, 2011; Qu *et al.*, 2017; Marcucci *et al.*, 2014) and failed to find the same prognostic meaning as published in three of them (Kelly *et al.*, 2017; Jost *et al.*, 2014; Hájková *et al.*, 2014). In case of Jost *et al.* (2014), we did not investigate the exact same locus because the particular CpG was not available in our data which can be the reason we did not confirm the reported prognostic ability of the aberrant DNA methylation. Sadly, we did not verify the results from our own previous studies. The methodological problems may be to blame in case of Hájková *et al.* (2012) where qMSP

was used. In case of Hájková *et al.* (2014) and Šestáková *et al.* (2019), the reasons may be an insufficiently large test cohort, only 79 AML patients, or the lack of an independent validation cohort, respectively.

There is no association between the studies and the regions for which we successfully validated the reported prognostic effect. Therefore, we provide only a summary of authors' discoveries. Božić *et al.* (2015) found a hypermethylation in *C1R* gene linked to poor prognosis of AML patients. *C1R* encodes a subcomponent of a complement protein involved in the innate immune system (GeneCards, 2020). The authors reported only a moderate correlation with gene expression and assigned the aberrant DNA methylation to global chromosomal changes (Božić *et al.*, 2015). Qu *et al.* (2017) discovered that lower methylation at two CpG island shores, one of a transcription regulator LZTS2 and second of a nuclear receptor NR6A1 (GeneCards, 2020), is associated with inferior outcome in AML. Next study described that an increased methylation in distal promoter of *CEBPA* is associated with lower expression of the gene and better survival of AML patients (Lin *et al.*, 2011). Lin *et al.* (2011) stressed in their work that this implies to patients without *CEBPA* mutations, thus, proving that the methylation level at that particular region is an independent prognostic marker able to further refine the AML prognosis. Marcucci *et al.* (2014) correlated differentially methylated promoters and gene expression levels and identified a set of seven genes where higher methylation and downregulated transcription were associated with better survival of AML patients. Two of the genes, encoding molecular marker *CD34* and miRNA *MIR155HG*, were previously associated with aggressive AML (Raspadori *et al.*, 1997; Marcucci *et al.*, 2013). Four other genes from the set have been linked to and adverse phenotype of various other tumors (Marcucci *et al.*, 2014) but otherwise represent a heterogenous group of genes associated with G-protein signaling (*RHOC*, *F2RL1*), exocytosis of mastocytes (*SCRN1*), and embryogenesis (*FAM92A1*) (GeneCards, 2020). The last gene from the seven-gene set was a poorly characterized ATPase *VWA8* (GeneCards, 2020). Deneberg *et al.* (2017) also based their analysis on promoter methylation levels coupled with gene expression. The authors described eighteen genes where promoter hypermethylation was associated with better survival of AML patients. The only thing these genes have in common is that they are regulated by polycomb group proteins (Deneberg *et al.*, 2011). Otherwise it is a heterogenous group of proteins comprising potassium channels (*KCNA1/4/6*), various enzymes (*BHMT*, *CHTS13*, *ETNK2*, *GAL3ST3*), proteins involved in intra- and extracellular trafficking (*MGC39715*, *SLC4A11*, *TF*), proteins regulating embryogenesis or cell development (*NEUROG1*, *SNCAIP*, *TCF15*), a protein

involved in apoptosis (*MEGF10*), and others with various functions (*COL21A1*, *CTS11*, *ESM1*, *OXGR1*) (GeneCards, 2020).

In general, the low number of regions for which we confirmed the prognostic impact with our NGS-based approach is not surprising. It rather highlights the importance of such validation studies and a need for a consistent and easy-to-reproduce approach to assess the impact of various changes in DNA methylation on AML prognosis.

5.3.2 MethScore as a new surrogate marker for AML

In this work, we introduced MethScore, a simply computed value that coherently evaluates the prognostic impact of various changes in DNA methylation. Our computation was inspired by the work of Marcucci *et al.* (2014). They used a similar method to count a summarizing score of differential gene expression.

The first step in MethScore assessment was to find CpGs with methylation levels associated with patients' survival. We identified a set of almost two thousand loci. In nearly half of them, lower methylation was linked to a better survival and, opposingly, to a worse outcome in the remaining CpGs. This is in accordance with the related literature where half of the studies reported adverse outcome of AML patients with hypermethylation and the other half with hypomethylation at specific regions. In the average methylation levels comparison of these selected CpGs, we did not see any striking difference in DNA methylation between the healthy donors' samples and the two AML cohorts. This is not unexpected given the dichotomous nature of the prognostically significant epigenetic changes. This similarity in global methylation levels was also reflected in the MethScore of healthy donors that was comparable with the average MethScore of AML patients.

From the figures showing MethScore values together with average methylation and number of mutations of individual patients, it seemed that higher MethScore slightly correlates with an increased number of mutations and lower overall methylation. The higher mutational burden may represent a progressing genome instability that is also characterized by substantial DNA methylation changes (Cai & Levine, 2019). The lower average methylation in patients with higher MethScore and thus adverse outcome may reflect the previously published discoveries that increased methylation at specific loci may serve as a break preventing AML progression (Spencer *et al.*, 2016) and thus higher DNA methylation is prognostically more favorable (Kroeger *et al.*, 2007; Hájková *et al.*, 2012).

In the Kaplan-Meier analyses of the test cohort, MethScore had a striking significance for both overall and event-free survivals which was further confirmed in a multivariate

Cox analysis. In the multivariate analyses, hazard ratios of mutations in *DNMT3A*, *ASXL1*, and *RUNX1* were opposite to what was expected, pointing to a favorable prognosis of patients with the particular mutation. In case of *DNMT3A*^{mut}, it may be caused by the co-mutational background of the patients. As already mentioned in the results section, a study by Gale *et al.* (2015) highlighted the effect of *NPM1* and *FLT3* co-mutations on the prognosis of patients with mutated *DNMT3A*. However, regarding *ASXL1* and *RUNX1* mutations, we do not have any satisfactory explanation. We evaluated the effect of those mutations separately in univariate analyses and there the hazard ratio pointed to the mutations having correctly an adverse effect on patients' survival. We also checked the proportionality assumption for each multivariate model to make sure of its accuracy. Thus, the discrepancy was probably caused only by the limited number of patients with these mutations, nine in both cases, that could not be properly statistically evaluated in the all-embracing regression model.

We further verified the predictive ability of MethScore on an independent validation cohort. The difference in Kaplan-Meier curves of patients with higher and lower MethScore was somewhat lower than in the test cohort, nevertheless, still statistically significant. This was again confirmed in multivariate Cox regression analyses. The difference between the two AML cohorts may be explained partially by the lower number of samples in the validation cohort. Moreover, there was a difference between the mutational backgrounds of each cohort mainly in the numbers of *NRAS*, *IDH2*, and *IDH1* mutations. Even though the total percentage of patients with *IDH1/2* mutations was similar in both cohorts (19 and 14 %), there was markedly fewer *IDH2*^{mut} patients in the validation cohort. This may have influenced the epigenetic landscape of the samples in accordance with the theory of diverse biological nature of the two mutations (Wang *et al.*, 2016) discussed in our first project and could introduce a bias into MethScore values.

In the comparison of patients with higher and lower MethScore, there was a significant difference in some of the aspects commonly used for the risk stratification of AML. It was age, cytogenetics, and *TP53* mutation in the test cohort and *FLT3-ITD* in the validation cohort. There were also more transplanted patients in the group with lower MethScore and thus better survival. This would imply that MethScore is just another surrogate marker further confirming the outcome of patients as already established with the traditional markers. Therefore, we focused only on patients with intermediate risk, that would benefit most from a better stratification (Döhner *et al.*, 2017). Patients were again perfectly stratified by MethScore in all the intermediate risk subgroups and the prognostic ability of MethScore

was again confirmed with multivariate analyses. This assured us of the validity and clinical applicability of MethScore.

5.3.3 The role of DNA methylation changes associated with *HOX* genes

The indispensable role of homeobox genes in hematopoiesis control is well-known and their impaired expression and aberrant DNA methylation have been implicated as a prognostic marker in AML (Deneberg *et al.*, 2011; Hájková *et al.*, 2012; Jost *et al.*, 2014). Therefore, when designing our DNA methylation panel, we included sequences covering all *HOX* genes.

Most of the CpGs from the set of loci with methylation significantly affecting the patient's survival were associated with *HOXA* and *HOXB* gene clusters. These genes are predominantly expressed in myeloid and erythroid cells (Alharbi *et al.*, 2012) and their upregulated expression is a negative prognostic marker in AML (Drabkin *et al.*, 2002; Chen *et al.*, 2019; Nagy *et al.*, 2019). Spencer *et al.* (2015) also showed that the overexpression of *HOXA* and *HOXB* genes is connected with their hypomethylation. Similar observation was reported by Jung *et al.* (2015) who described overexpression and hypomethylation of *HOXA* genes as a key feature of leukemia stem cells signature connected with worse survival of AML patients. Within *HOXA* CpGs significantly associated with survival, those with lower methylation related to better outcome prevailed which would seem contradictory to the literature. However, a closer look at the methylation levels of AML patients with comparison to healthy donors revealed a distinct region hypomethylated in patients with short survival. This region overlaps with a 38 kbp region reported as regulatory for *HOXA* genes (Spencer *et al.*, 2015). Similarly, we noted a hypomethylation in worse surviving patients in regulatory region of *HOXB* cluster overlapping a ~ 50 bp *HOXB* control region also described by Spencer *et al.* (2015).

HOXC genes are typically expressed in lymphoid cells (Alharbi *et al.*, 2012) and thus it is not too surprising that such a small number of CpGs significant for AML patient's survival was associated with this gene cluster. Interestingly, *HOXD* genes, which are not expressed during hematopoiesis (Alharbi *et al.*, 2012), were quite highly represented. Aberrant methylation of *HOXD* genes in AML has been reported (Jelinek *et al.*, 2008; Hájková *et al.*, 2012). Unfortunately, the authors did not mention any implications for patients' prognosis. Generally, we can presume that the detected favorable hypermethylation of these genes indicates their silencing based on the consistently reported inverse correlation between expression and methylation of *HOX* genes (Strathdee *et al.*, 2007; Jelinek *et al.*, 2008; Di Vinci *et al.*, 2011).

The regulation processes of *HOX* gene clusters are very complex, requiring tight cooperation between DNA methylation and histone modifications. Also, two other homeobox genes, *PBX* and *MEIS*, are known cofactors recruited by *HOX* genes for specific DNA binding (Moens & Selleri, 2006). In our set of CpGs with methylation significant for patients' outcome, we also had loci linked to *PBX3* (n = 16), *MEIS1* (n = 4) and *MEIS1-AS* (n = 44), which may further affect the *HOX*-related processes in AML progression. Moreover, the mutational background of patients can also influence the *HOX* gene expression, especially the *MLL* translocations and *NPM1* mutations (He *et al.*, 2011; Alharbi *et al.*, 2012). Therefore, without more thorough research we are not able to properly estimate the real biological role of DNA methylation changes associated with *HOX* genes in AML pathogenesis. But still, our data showed distinct regions within *HOX* gene clusters that may be relevant for further investigation.

6 CONCLUSIONS

Our results showed that distinct mutational backgrounds of AML patients based on mutations in *DNMT3A* and *IDH1/2* are characterized by specific DNA methylation and hydroxymethylation profiles. We observed a mixed DNA hydroxy-/methylation profile in patients with concurrent mutations in both *DNMT3A* and *IHD1/2* genes. Detected epigenetic alterations were not entirely reflected by changes in the gene expression. Furthermore, there was a considerable difference in numbers and sites with aberrant DNA hydroxy-/methylation and also in differentially expressed genes between $IHD1^{mut}$ and $IDH2^{mut}$ patients which supports the theory that these mutations represent distinct biological entities.

We observed in the array data and confirmed with the mass spectrometry analysis that levels of DNA hydroxymethylation in blood samples are very low in comparison with DNA methylation. Thus, we assume that the role of DNA hydroxymethylation in AML pathogenesis is not as important as that of DNA methylation.

We detected a hypermethylation in an upstream enhancer linked to *GZMB* gene, encoding a mediator of apoptosis, that was associated with an inferior survival of AML patients. Unfortunately, we were not able to determine the biological mechanism behind this altered DNA methylation.

We compared four standard approaches for validation of DNA methylation in order to find the most appropriate method that would be suitable for a common laboratory practice. Pyrosequencing together with MS-HRM were rated the best in terms of overall feasibility and results' consistency across different methylation levels. The base resolution of pyrosequencing represented the main advantage over the other approaches. MSRE analysis proved to be the fastest method but not suitable for precise evaluation of intermediately methylated regions. qMSP approach was the most demanding and did not provide satisfactory results.

We introduced a DNA methylation sequencing panel targeting sites with aberrant DNA methylation that were described as prognostically significant for AML patients in twenty publications. With our data, we were able to successfully validate the reported significance for patient's survival in five of those studies. We did not verify the prognostic relevance of more regions probably due to the different methodologies used for the assessment of DNA methylation levels and their significance or an insufficient number of tested samples, in some cases without a validation cohort in the original publications. Our results underline

the importance of independent validation studies that are essential for the translation of DNA methylation changes into clinical decision making.

For a comprehensive evaluation of all investigated DNA methylation changes, we developed a summarizing value and called it MethScore. MethScore stratified with high accuracy patients with better and worse survival and its prognostic significance was confirmed in multivariate analyses. We validated the ability of MethScore to separate patients with longer and shorter survival on an independent cohort of AML patients. Furthermore, we showed that MethScore is able to reliably distinguish also patients within the intermediate risk group that may mostly benefit from a better stratification.

From our data, it is apparent that aberrant methylation of *HOX* genes affects the outcome of AML patients. We discovered DNA hypomethylation in regulatory regions of *HOXA* and *HOXB* clusters in patients with short survival. We also observed a hypermethylation of *HOXC* and *HOXD* genes favorable for patients' outcome. However, the regulation processes of *HOX* genes are extremely complex and will require a thorough investigation to draw more definite conclusions.

7 LIST OF ABBREVIATIONS

5-hmC	5-hydroxymethylcytosine
5-mC	5-methylcytosine
ABL1	Abelson Tyrosine-Protein Kinase 1
AC012531.2	lncRNA associated with HOXC cluster
AKT	AKT Serine/Threonine Kinase
AML	Acute myeloid leukemia
APL	Acute promyelocytic leukemia
ASXL1/2	Additional Sex Combs Like 1/2, Transcriptional Regulator
ATO	Arsenic trioxide
ATRA	All-*trans* retinoic acid
AUC	Area under the curve
BCR	Breakpoint Cluster Region Protein
BM	Bone marrow
bp	Base pairs
BS	Bisulfite
C1R	complement component 1 subcomponent R
CALCA	Calcitonin Related Polypeptide Alpha
CBF	Core binding factor
CBF-AML	Core binding factor acute myeloid leukemia
CBFB	Core-Binding Factor Subunit Beta
CBL	Cbl Proto-Oncogene
CD34	Hematopoietic Progenitor Cell Antigen CD34
CDH1	Cadherin 1
CDKN2B	Cyclin Dependent Kinase Inhibitor 2B
CEBPA	CCAAT/enhancer-binding protein alpha
$CEBPA^{dm}$	CEBPA double mutant
CGI	CpG islands
CI	Confidence interval
CIMP	CGI methylator phenotype
CN-AML	Cytogenetically normal acute myeloid leukemia
CpG	CG dinucleotide
CR	Complete remission
C_t	Threshold cycle
DEK	DEK Proto-Oncogene
DLX4	Distal-Less Homeobox 4
DMP	Differentially methylated position
DMR	Differentially methylated region
DNMT	DNA methyltransferase
ELN	European Leukemia Net
ESR1	Estrogen Receptor 1
EZH2	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit
F2RL1	F2R Like Trypsin Receptor 1

FAB	French-American-British AML classification
FAM92A1	Family With Sequence Similarity 92 Member A1
FDA	U.S. Food and Drug Administration
FLT3	FMS-like tyrosine kinase 3
FLT3-ITD	FLT3 kinase with internal tandem duplications
FLT3-TKD	FLT3 kinase with mutated tyrosine kinase domain
GATA2	GATA Binding Protein 2
GPX3	Glutathione Peroxidase 3
GZMB	Granzyme B
HELP	HpaII tiny fragment enrichment by ligation-mediated PCR
HMAs	Hypomethylating agents
hmCpGs	Hydroxymethylated positions
HOTTIP	lncRNA associated with HOXA cluster
HOX	Homeobox
HR	Hazard ratio
HRM	High resolution melting
HSC	Hematopoietic stem cells
CHFR	Checkpoint With Forkhead And Ring Finger Domains
IDH1/2	Isocitrate dehydrogenase
IM	Intermediately methylated locus/region
IR	Patients from intermediate cytogenetics risk group
IR^{FLT3+}	Patients with intermediate cytogenetics and *FLT3*-ITD
$IR^{FLT3+NPM1+}$	Patients with intermediate cytogenetics and *FLT3*-ITD and $NPM1^{mut}$
IR^{NPM1-}	Patients with intermediate cytogenetics and no mutation in *NPM1* gene
KIT	KIT Proto-Oncogene, Receptor Tyrosine Kinase
KMT2A (MLL)	Lysine Methyltransferase 2A
LC	Liquid chromatography
lncRNA	Long non-coding RNA
LZTS2	Leucine Zipper Tumor Suppressor 2
M	Methylated locus/region
MBD	5-methylcytosine binding protein
MBP	Methyl-CpG binding proteins
MDS	Myelodysplastic syndrome
MECOM	MDS1 And EVI1 Complex Locus
MeDIP	Methylated DNA immunoprecipitation
MEIS	Myeloid Ecotropic Viral Integration Site 1 Homolog
MEK	Mitogen-Activated Protein Kinase Kinase 1
MELP	Microsphere HpaII tiny fragment enrichment by ligation-mediated PCR
MIR155HG	MIR155 Host Gene
miRNA	Micro RNA
MLL	Lysine Methyltransferase 2A
MNC	Mononuclear cells
MRD	Minimal residual disease

mRNA	Messenger RNA
MS	MethScore
MS/MS	Tandem mass spectrometry
MS-HRM	Methylation specific high-resolution melting analysis
MSP Met	Methylated methylation-specific primers
MSP Unm	Unmethylated methylation-specific primers
MS-PCR	Methylation specific PCR
MSRE	Methylation-specific restriction enzyme
MYB	MYB Proto-Oncogene, Transcription Factor
MYH11	Myosin Heavy Chain 11
MYOD1	Myogenic Differentiation 1
NA	Not analyzed
NF1	Neurofibromin 1
NGS	Next-Generation Sequencing
NK cells	Natural killer cells
NKD2	NKD Inhibitor Of WNT Signaling Pathway 2
NPM1	Nucleophosmin
NPM2	Nucleophosmin/Nucleoplasmin 2
NR6A1	Nuclear Receptor Subfamily 6 Group A Member 1
NRAS	Neuroblastoma RAS Viral (V-Ras) Proto-Oncogene, GTPase
NUP214	Nucleoporin 214
OS	Overall survival
OSCP1	Organic Solute Carrier Partner 1
PB	Peripheral blood
PBX	Pre-B-Cell Leukemia Homeobox
PcG	Polycomb group proteins
PCR	Polymerase chain reaction
PI3K	Phosphoinositide 3-kinase
PML	PML Nuclear Body Scaffold
PTPN11	Protein Tyrosine Phosphatase Non-Receptor Type 11
PU.1 (SPI1)	Spi-1 Proto-Oncogene
qMSP	Quantitative methylation-specific PCR
RAD21	RAD21 Cohesin Complex Component
RAF	Raf-1 Proto-Oncogene, Serine/Threonine Kinase
RARA	Retinoic Acid Receptor Alpha
RASSF1A	Ras Association Domain Family Member 1
RHOC	Ras Homolog Family Member C
RNF216	Ring Finger Protein 216
RRBS	Reduced representation bisulfite sequencing
RUNX1	Runt-Related Transcription Factor 1
RUNX1T1	RUNX1 Partner Transcriptional Co-Repressor 1
SCGB3A1	Secretoglobin Family 3A Member 1
SCRN1	Secernin 1
SD	Standard deviation

SF3B1	Splicing Factor 3b Subunit 1
SFRP1/2	Secreted Frizzled Related Protein 1/2
siRNA	Small interfering RNA
SMC1A	Structural Maintenance Of Chromosomes 1A
SOCS1/2	Suppressor Of Cytokine Signaling 1/2
SRSF2	Serine And Arginine Rich Splicing Factor 2
STAG	Stromal Antigen
TAF1	TATA-Box Binding Protein Associated Factor 1
T_{ann}	Annealing temperature
TCGA	The cancer genome atlas
TET2	ten-eleven translocation oncogene family member 2
T_m	Melting temperature
TP53	Tumor Protein P53
U	Unmethylated locus/region
U2AF1	U2 Small Nuclear RNA Auxiliary Factor 1
VAF	Variant allele frequency
VTRNA2-1	Vault RNA 2-1
VWA8	Von Willebrand Factor A Domain Containing 8
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organization
WT1	Wilms Tumor 1, Transcription Factor
xMELP	Expedited microsphere HpaII small fragment enrichment by ligation-mediated PCR

8 REFERENCES

Abdel-Wahab O, Levine RL. **Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia.** Blood 2013;**121**(18):3563-72.

Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttil J, Zhang L, Khrebtukova I et al. **Base-Pair Resolution DNA Methylation Sequencing Reveals Profoundly Divergent Epigenetic Landscapes in Acute Myeloid Leukemia.** PLoS genetics 2012;**8**(6):e1002781.

Al Harrasi I, Al-Yahyai R, Yaish MW. **Detection of Differential DNA Methylation Under Stress Conditions Using Bisulfite Sequence Analysis.** Methods in molecular biology (Clifton, N.J.) 2017;**1631**:121-37.

Alharbi RA, Pettengell R, Pandha HS, Morgan R. **The role of HOX genes in normal hematopoiesis and acute leukemia.** Leukemia 2012;**27**(5):1000-8.

Andrews S. **FastQC: a quality control tool for high throughput sequence data.** Babraham Bioinformatics 2010;.

Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM et al. **The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia.** Blood 2016;**127**(20):2391-405.

Bair E, Hastie T, Paul D, Tibshirani R. **Prediction by Supervised Principal Components.** Journal of the American Statistical Association 2006;**101**(473):119-37.

Behdad A, Betz BL. **Molecular Testing in Acute Myeloid Leukemia.** In: Coleman WB, Tsongalis GJ, editors. Diagnostic Molecular Pathology: A Guide to Applied Molecular Testing: Academic Press; 2016, p. 419-434.

Bennet JM, Catovsky D, Daniel M, Flandrin G, Galton D, Gralnick Ht et al. **Proposals for the classification of the acute leukemias.** Br J Haematol 1976;**33**(4):451-8.

Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W et al. **Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution.** Science (American Association for the Advancement of Science) 2012;**336**(6083):934-7.

Božić T, Lin Q, Frobel J, Wilop S, Hoffmann M, Müller-Tidow C et al. **DNA-methylation in C1R is a prognostic biomarker for acute myeloid leukemia.** Clinical epigenetics 2015;**7**(1):116.

Cai SF, Levine RL. **Genetic and epigenetic determinants of AML pathogenesis.** Seminars in Hematology 2019;**56**(2):84-9.

Cao X, Cai SF, Fehniger TA, Song J, Collins LI, Piwnica-Worms DR et al. **Granzyme B and Perforin Are Important for Regulatory T Cell-Mediated Suppression of Tumor Clearance.** Immunity (Cambridge, Mass.) 2007;**27**(4):635-46.

Chen S, Qin Z, Hu F, Wang Y, Dai Y, Liang Y. **The Role of the HOXA Gene Family in Acute Myeloid Leukemia.** Genes 2019;**10**(8):621.

Chiusolo V, Jacquemin G, Yonca Bassoy E, Vinet L, Liguori L, Walch M et al. **Granzyme B enters the mitochondria in a Sam50-, Tim22- and mtHsp70-dependent manner to induce apoptosis.** Cell death and differentiation 2017;**24**(4):747-58.

Corces-Zimmerman MR, Hong W, Weissman IL, Medeiros BC, Majeti R. **Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission.** Proceedings of the National Academy of Sciences of the United States of America 2014;**111**(7):2548-53.

Curik N, Burda P, Vargova K, Pospisil V, Belickova M, Vlckova P et al. **5-Azacitidine in aggressive myelodysplastic syndromes regulates chromatin structure at PU.1 gene and cell differentiation capacity.** Leukemia 2012;**26**(8):1804-11.

Delaney C, Garg SK, Yung R. **Analysis of DNA methylation by pyrosequencing.** Methods in molecular biology (Clifton, N.J.) 2015;**1343**:249.

Deneberg S, Guardiola P, Lennartsson A, Qu Y, Gaidzik V, Blanchet O et al. **Prognostic DNA methylation patterns in cytogenetically normal acute myeloid leukemia are predefined by stem cell chromatin marks.** Blood 2011;**118**(20):5573-82.

Di Vinci A, Casciano I, Marasco E, Banelli B, Ravetti GL, Borzì L et al. **Quantitative methylation analysis of HOXA3, 7, 9, and 10 genes in glioma: association with tumor WHO grade and clinical outcome.** J Cancer Res Clin Oncol 2011;**138**(1):35-47.

Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T et al. **Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel.** Blood 2017;**129**(4):424-47.

Döhner H, Weisdorf DJ, Bloomfield CD. **Acute Myeloid Leukemia.** The New England Journal of Medicine 2015;**373**(12):1136-52.

Doubek M, Mayer J. **Postupy diagnostiky a léčby leukemií a jejich infekčních komplikací u dospělých pacientů.** , 2013.

Drabkin H, Parsy C, Ferguson K, Guilhot F, Lacotte L, Roy L et al. **Quantitative HOX expression in chromosomally defined subsets of acute myelogenous leukemia.** Leukemia 2002;**16**(2):186-95.

Duployez N, Marceau-Renaut A, Boissel N, Petit A, Bucci M, Geffroy S et al. **Comprehensive mutational profiling of core binding factor acute myeloid leukemia.** Blood 2016;**127**(20):2451-9.

Dwight Z, Palais R, Wittwer C. **uAnalyze: web-based high-resolution DNA melting analysis with comparison to thermodynamic predictions.** TCBB 2012;**9**(6):1805-11.

Eads CA, Danenberg KD, Kawakami K, Saltz LB et al. **MethyLight: a high-throughput assay to measure DNA methylation.** Nucleic Acids Res 2000;**28**(8):e32.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. **GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.** BMC bioinformatics 2009;**10**(1):48.

Ewels P, Magnusson M, Lundin S, Käller M. **MultiQC: summarize analysis results for multiple tools and samples in a single report.** Computer applications in the biosciences 2016;**32**(19):3047-8.

Falini B, Mecucci C, Tiacci E, Alcalay M, Rosati R, Pasqualucci L et al. **Cytoplasmic Nucleophosmin in Acute Myelogenous Leukemia with a Normal Karyotype.** The New England Journal of Medicine 2005;**352**(3):254-66.

Fasan A, Haferlach C, Alpermann T, Jeromin S, Grossmann V, Eder C et al. **The role of different genetic subtypes of CEBPA mutated AML.** Leukemia 2014;**28**(4):794-803.

Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A et al. **Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation.** Cancer cell 2010;**18**(6):553-67.

Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ et al. **DNA Methylation Signatures Identify Biologically Distinct Subtypes in Acute Myeloid Leukemia.** Cancer cell 2010;**17**(1):13-27.

Fisher CL, Randazzo F, Humphries RK, Brock HW. **Characterization of Asxl1, a murine homolog of Additional sex combs, and analysis of the Asx-like gene family.** Gene 2006;**369**:109-18.

Folta A, Culen M, Jeziskova I, Herudkova Z, Tom N, Hlubinkova T et al. **Prognostic significance of mutation profile at diagnosis and mutation persistence during disease remission in adult acute myeloid leukaemia patients.** British Journal of Haematology 2019;**186**(2):300-10.

Fong CY, Morison J, Dawson MA. **Epigenetics in the hematologic malignancies.** Haematologica 2014;**99**(12):1772-83.

Fortin J, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ et al. **Functional normalization of 450k methylation array data improves replication in large cancer studies.** Genome biology 2014;**15**(11):503.

Frommer M, Mcdonald LE, Millar DS, Collis CM et al. **A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.** 1992;.

Gaidzik VI, Teleanu V, Papaemmanuil E, Weber D, Paschka P, Hahn J et al. **RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features.** Leukemia 2016;**30**(11):2282.

Gale RE, Lamb K, Allen C, El-Sharkawi D, Stowe C, Jenkinson S et al. **Simpson's Paradox and the Impact of Different DNMT3A Mutations on Outcome in Younger Adults With Acute Myeloid Leukemia.** Journal of clinical oncology 2015;**33**(18):2072-83.

Gao L, Liu F, Zhang H, Sun J, Ma Y. **CHFR hypermethylation, a frequent event in acute myeloid leukemia, is independently associated with an adverse outcome.** Genes, Chromosomes and Cancer 2016;**55**(2):158-68.

Garzon R, Heaphy CEA, Havelange V, Fabbri M, Volinia S, Tsao T et al. **MicroRNA 29b functions in acute myeloid leukemia.** Blood 2009;**114**(26):5331-41.

Garzon R, Volinia S, Liu C, Fernandez-Cymering C, Palumbo T, Pichiorri F et al. **MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia.** Blood 2008;**111**(6):3183-9.

Garzon R, Volinia S, Papaioannou D, Nicolet D, Kohlschmidt J, Yan PS et al. **Expression and prognostic impact of lncRNAs in acute myeloid leukemia.** Proceedings of the National Academy of Sciences - PNAS 2014;**111**(52):18679-84.

Gebhard C, Benner C, Ehrich M, Schwarzfischer L, Schilling E, Klug M et al. **General Transcription Factor Binding at CpG Islands in Normal Cells Correlates with Resistance to De novo DNA Methylation in Cancer Cells.** Cancer Research 2010;**70**(4):1398-407.

Gebhard C, Glatz D, Schwarzfischer L, Wimmer J, Stasik S, Nuetzel M et al. **Profiling of aberrant DNA methylation in acute myeloid leukemia reveals subclasses of CG-rich regions with epigenetic or genetic association.** Leukemia 2019;**33**(1):26-36.

GeneCards. **GeneCards®: The Human Gene Database.** www.genecards.org 2020.

Glass JL, Hassane D, Wouters BJ, Kunimoto H, Avellino R, Garrett-Bakelman FE et al. **Epigenetic Identity in AML Depends on Disruption of Nonpromoter Regulatory Elements and Is Affected by Antagonistic Effects of Mutations in Epigenetic Modifiers.** Cancer discovery 2017;**7**(8):868-83.

Grafone T, Palmisano M, Nicci C, Storti S. **An overview on the role of FLT3-tyrosine kinase receptor in acute myeloid leukemia: biology and treatment.** Oncology reviews 2012;**6**(1):8-e8.

Grimwade D, Walker H, Harrison G, Oliver F, Chatters S, Harrison CJ et al. **The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial.** Blood 2001;**98**(5):1312-20.

Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH et al. **Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials.** Blood 2010;**116**(3):354-65.

Grimwade D, Ivey A, Huntly BJP. **Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance.** Blood 2016;**127**(1):29-41.

Grimwade D, Mrózek K. **Diagnostic and Prognostic Value of Cytogenetics in Acute Myeloid Leukemia.** Hematology/Oncology Clinics of North America 2011;**25**(6):1135-61.

Guo H, Zhang T, Wen X, Zhou J, Ma J, An C et al. **Hypermethylation of secreted frizzled-related proteins predicts poor prognosis in non-M3 acute myeloid leukemia.** OncoTargets and therapy 2017;**10**:3635-44.

Guryanova OA, Shank K, Spitzer B, Luciani L, Koche RP, Garrett-Bakelman FE et al. **DNMT3A mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling.** Nature medicine 2016;**22**(12):1488-95.

Haferlach C, Dicker F, Herholz H, Schnittger S, Kern W, Haferlach T. **Mutations of the TP53 gene in acute myeloid leukemia are strongly associated with a complex aberrant karyotype.** Leukemia 2008;**22**(8):1539-41.

Hájková H. **DNA Methylation Changes in Patients with Acute Myeloid Leukemia.** 2015;.

Hájková H, Fritz MH, Haškovec C, Schwarz J, Šálek C, Marková J et al. **CBFB-MYH11 hypomethylation signature and PBX3 differential methylation revealed by targeted bisulfite sequencing in patients with acute myeloid leukemia.** Journal of hematology & oncology 2014;**7**(1):66.

Hájková H, Marková J, Haškovec C, Šárová I, Fuchs O, Kostečka A et al. **Decreased DNA methylation in acute myeloid leukemia patients with DNMT3A mutations and prognostic implications of DNA methylation.** Leukemia Research 2012;**36**(9):1128-33.

Hansen LL, Wojdacz TK, Dobrovic A. **Methylation-sensitive high-resolution melting.** Nature Protocols 2008;**3**(12):1903-8.

Hayatsu H, Wataya Y, Kazushige K. **Addition of sodium bisulfite to uracil and to cytosine.** Journal of the American Chemical Society 1970;**92**(3):724-6.

He H, Hua X, Yan J. **Epigenetic regulations in hematopoietic Hox code.** Oncogene 2011;**30**(4):379-88.

He X, Zhu Y, Lin Y, Li M, Du J, Dong H et al. **PRMT1-mediated FLT3 arginine methylation promotes maintenance of FLT3-ITD+ acute myeloid leukemia.** Blood 2019;**134**(6):548-60.

Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB. **Methylation-Specific PCR: A Novel PCR Assay for Methylation Status of CpG Islands.** Proceedings of the National Academy of Sciences of the United States of America 1996;**93**(18):9821-6.

Husseiny MI, Kuroda A, Kaye AN, Nair I, Kandeel F, Ferreri K. **Development of a Quantitative Methylation-Specific Polymerase Chain Reaction Method for Monitoring Beta Cell Death in Type 1 Diabetes.** PLoS One 2012;**7**(10):e47942.

Itoi E, Kokubun S, Hashimoto K, Roach HI. **Improved Quantification of DNA Methylation Using Methylation-Sensitive Restriction Enzymes and Real-Time PCR.** Epigenetics 2007;**2**(2):86-91.

Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG et al. **Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes.** The New England Journal of Medicine 2014;**371**(26):2488-98.

Jelinek J, Kroeger H, He R, Saraf A, Lau SF, Nguyen T et al. **Hypermethylation of HOX Genes Is Associated with Longer Survival in Acute Myeloid Leukemia.** Blood 2008;**112**(11):312.

Jeschke J, Collignon E, Fuks F. **Portraits of TET-mediated DNA hydroxymethylation in cancer.** Current opinion in genetics & development 2016;**36**:16-26.

Johnson WE, Li C, Rabinovic A. **Adjusting batch effects in microarray expression data using empirical Bayes methods.** Biostatistics (Oxford, England) 2007;**8**(1):118-27.

Jones PA. **Functions of DNA methylation: islands, start sites, gene bodies and beyond.** Nature reviews. Genetics 2012;**13**(7):484-92.

Jost E, Lin Q, Weidner CI, Wilop S, Hoffmann M, Walenda T et al. **Epimutations mimic genomic mutations of DNMT3A in acute myeloid leukemia.** Leukemia 2014;**28**(6):1227-34.

Jourdan E, Boissel N, Chevret S, Delabesse E, Renneville A, Cornillet P et al. **Prospective evaluation of gene mutations and minimal residual disease in patients with core binding factor acute myeloid leukemia.** Blood 2013;**121**(12):2213-23.

Jung N, Dai B, Gentles AJ, Majeti R, Feinberg AP. **An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis.** Nature communications 2015;**6**(1):8489.

Kadia TM, Jain P, Ravandi F, Garcia-Manero G, Andreef M, Takahashi K et al. **TP53 mutations in newly diagnosed Acute Myeloid Leukemia –Clinico-molecular characteristics, response to therapy, and outcomes.** Cancer 2016;**122**(22):3484-91.

Kelly AD, Kroeger H, Yamazaki J, Taby R, Neumann F, Yu S et al. **A CpG island methylator phenotype in acute myeloid leukemia independent of IDH mutations and associated with a favorable outcome.** Leukemia 2017;**31**(10):2011-9.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM et al. **The Human Genome Browser at UCSC.** Genome research 2002;**12**(6):996-1006.

Khulan B. **Comparative isoschizomer profiling of cytosine methylation: The HELP assay.** Genome research 2006;**16**(8):1046-55.

Ko M, An J, Rao A. **DNA methylation and hydroxymethylation in hematologic differentiation and transformation.** Current opinion in cell biology 2015;**37**:91-101.

Kroeger H, Jelinek J, Kornblau SM, Bueso-Ramos CE, Issa J. **Increased DNA Methylation Is Associated with Good Prognosis in AML.** Blood 2007;**110**(11):595.

Kroeze L, Aslanyan MG, Rooij Av, Koorenhof-Scheele TN, Massop M, Carell T et al. **Characterization of acute myeloid leukemia based on levels of global hydroxymethylation.** Blood 2014;**124**(7):1110-8.

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z et al. **Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.** Nucleic acids research 2016;**44**(W1):W90-7.

Kurdyukov S, Bullock M. **DNA Methylation Analysis: Choosing the Right Method.** Biology (Basel, Switzerland) 2016;**5**(1):3.

Le T, Kim K, Fan G, Faull KF. **A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples.** Analytical Biochemistry 2011;**412**(2):203-9.

Lei L, Xia S, Liu D, Li X, Feng J, Zhu Y et al. **Genome-wide characterization of lncRNAs in acute myeloid leukemia.** Briefings in bioinformatics 2018;**19**(4):627-35.

Leisch M, Jansko B, Zaborsky N, Greil R, Pleyer L. **Next Generation Sequencing in AML-On the Way to Becoming a New Standard for Treatment Initiation and/or Modulation?** Cancers 2019;**11**(2):252.

Levine R, Valk P. **Next-generation sequencing in the diagnosis and minimal residual disease assessment of acute myeloid leukemia.** Haematologica 2019;**104**(5):868-71.

Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE et al. **DNMT3A Mutations in Acute Myeloid Leukemia.** The New England Journal of Medicine 2010;**363**(25):2424-33.

Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson AG et al. **Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia.** The New England Journal of Medicine 2013;**368**(22):2059-74.

Li S, Garrett-Bakelman F, Chung SS, Sanders M, Hricik T, Rapaport F et al. **Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia.** Nature Medicine 2016;**22**(7):792-9.

Li X, Zhou J, Zhang T, Yang L, Wen X, Ma J et al. **Epigenetic dysregulation of NKD2 is a valuable predictor assessing treatment outcome in acute myeloid leukemia.** Journal of Cancer 2017;**8**(3):460-8.

Lin T, Hou H, Chou W, Ou D, Yu S, Tien H et al. **CEBPA methylation as a prognostic biomarker in patients with de novo acute myeloid leukemia.** Leukemia 2011;**25**(1):32-40.

Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL et al. **Acute myeloid leukemia ontogeny is defined by distinct somatic mutations.** Blood 2015;**125**(9):1367-76.

Lindström MS. **NPM1/B23: A Multifunctional Chaperone in Ribosome Biogenesis and Chromatin Remodeling.** Biochemistry research international 2011;**2011**:195209-16.

Liu F, Gong M, Gao L, Cai X, Zhang H, Ma Y. **RASSF1A hypermethylation is associated with ASXL1 mutation and indicates an adverse outcome in non-M3 acute myeloid leukemia.** OncoTargets and therapy 2017;**10**:4143-51.

Liu X, Gong Y. **Isocitrate dehydrogenase inhibitors in acute myeloid leukemia.** Biomarker research 2019;**7**(1):22.

Lunnon K, Hannon E, Smith RG, Dempster E, Wong C, Burrage J et al. **Variation in 5-hydroxymethylcytosine across human cortex and cerebellum.** Genome biology 2016;**17**(1):27.

Luskin MR, Gimotty PA, Smith C, Loren AW, Figueroa ME, Harrison J et al. **A clinical measure of DNA methylation predicts outcome in de novo acute myeloid leukemia.** JCI insight 2016;**1**(9).

Marcucci G, Maharry KS, Metzeler KH, Volinia S, Wu Y, Mrózek K et al. **Clinical Role of microRNAs in Cytogenetically Normal Acute Myeloid Leukemia: miR-155 Upregulation Independently Identifies High-Risk Patients.** Journal of Clinical Oncology 2013;**31**(17):2086-93.

Marcucci G, Yan P, Maharry K, Frankhouser D, Nicolet D, Metzeler KH et al. **Epigenetics Meets Genetics in Acute Myeloid Leukemia: Clinical Impact of a Novel Seven-Gene Score.** Journal of Clinical Oncology 2014;**32**(6):548-56.

Martin M. **Cutadapt removes adapter sequences from high-throughput sequencing reads.** EMBnet.journal 2011;**17**(1):10.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB et al. **GREAT improves functional interpretation of cis-regulatory regions.** Nature biotechnology 2010;**28**(5):495-501.

Meissner A. **Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis.** Nucleic acids research 2005;**33**(18):5868-77.

Metzeler KH, Herold T, Rothenberg-Thurley M, Amler S, Sauerland MC, Görlich D et al. **Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia.** Blood 2016;**128**(5):686-98.

Micol J, Duployez N, Boissel N, Petit A, Geffroy S, Nibourel O et al. **Frequent ASXL2 mutations in acute myeloid leukemia patients with t(8;21)/RUNX1-RUNX1T1 chromosomal translocations.** Blood 2014;**124**(9):1445-9.

Miyamoto T, Weissman IL, Akashi K. **AML1/ETO-Expressing Nonleukemic Stem Cells in Acute Myelogenous Leukemia with 8;21 Chromosomal Translocation.** Proceedings of the National Academy of Sciences of the United States of America 2000;**97**(13):7521-6.

Moens CB, Selleri L. **Hox cofactors in vertebrate development.** Developmental biology 2006;**291**(2):193-206.

Montalban-Bravo G, DiNardo CD. **The role of IDH mutations in acute myeloid leukemia.** Future Oncology 2018;**14**(10):979-93.

Nagy Á, Ősz Á, Budczies J, Krizsán S, Szombath G, Demeter J et al. **Elevated HOX gene expression in acute myeloid leukemia is associated with NPM1 mutations and poor survival.** Journal of advanced research 2019;**20**:105-16.

National Cancer Institute. **NIH: Surveillance, Epidemiology, and End Results (SEER) Program.** https://www.cancer.gov/ 2020.

Norvil AB, AlAbdi L, Liu B, Tu YH, Forstoffer NE, Michie AR et al. **The acute myeloid leukemia variant DNMT3A Arg882His is a DNMT3B-like enzyme.** Nucleic acids research 2020;**48**(7):3761-75.

Otto C, Stadler PF, Hoffmann S. **Fast and sensitive mapping of bisulfite-treated sequencing data.** Computer applications in the biosciences 2012;**28**(13):1698-704.

Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND et al. **Genomic Classification and Prognosis in Acute Myeloid Leukemia.** The New England Journal of Medicine 2016;**374**(23):2209-21.

Paschka P, Schlenk RF, Gaidzik VI, Herzig JK, Aulitzky T, Bullinger L et al. **ASXL1 mutations in younger adult patients with acute myeloid leukemia: a study by the German-Austrian Acute Myeloid Leukemia Study Group.** Haematologica (Roma) 2015;**100**(3):324-30.

Pfaffl MW. **Real-time qPCR amplification efficiency.** In: Dorak MT, editor. Real-time PCR, London: Taylor & Francis; 2007, p. 68-71.

Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P et al. **Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling.** Genome biology 2016;**17**(1):208.

Pratcorona M, Abbas S, Sanders MA, Koenders JE, Kavelaars FG, Erpelinck-Verschueren CAJ et al. **Acquired mutations in ASXL1 in acute myeloid leukemia: prevalence and prognostic value.** Haematologica 2012;**97**(3):388-92.

Qu X, Othus M, Davison J, Wu Y, Yan L, Meshinchi S et al. **Prognostic methylation markers for overall survival in cytogenetically normal patients with acute myeloid leukemia treated on SWOG trials.** Cancer 2017;**123**(13):2472-81.

Rampal R, Alkalin A, Madzo J, Vasanthakumar A, Pronier E, Patel J et al. **DNA Hydroxymethylation Profiling Reveals that WT1 Mutations Result in Loss of TET2 Function in Acute Myeloid Leukemia.** Cell reports (Cambridge) 2014;**9**(5):1841-55.

Raspadori D, Lauria F, Ventura MA, Rondelli D, Visani G, de Vivo A et al. **Incidence and prognostic relevance of CD34 expression in acute myeloblastic leukemia: Analysis of 141 cases.** Leukemia research 1997;**21**(7):603-7.

Reed K, Poulin ML, Yan L, Parissenti AM. **Comparison of bisulfite sequencing PCR with pyrosequencing for measuring differences in DNA methylation.** Analytical Biochemistry 2010;**397**(1):96-106.

Rücker FG, Schlenk RF, Bullinger L, Kayser S, Teleanu V, Kett H et al. **TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome.** Blood 2012;**119**(9):2114-21.

Russler-Germain D, Spencer D, Young M, Lamprecht T, Miller C, Fulton R et al. **The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by Blocking Its Ability to Form Active Tetramers.** Cancer Cell 2014;**25**(4):442-54.

Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J et al. **Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer.** Nature genetics 2006;**39**(2):232-6.

Schoofs T, Berdel WE, Müller-Tidow C. **Origins of aberrant DNA methylation in acute myeloid leukemia.** Leukemia 2014;**28**(1):1-14.

Schubeler D. **Function and information content of DNA methylation.** Nature 2015;**517**(7534):321.

Sedelies KA, Sayers TJ, Edwards KM, Chen W, Pellicci DG, Godfrey DI et al. **Discordant Regulation of Granzyme H and Granzyme B Expression in Human Lymphocytes.** Journal of Biological Chemistry 2004;**279**(25):26581-7.

Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V et al. **Identification of pre-leukemic hematopoietic stem cells in acute leukemia.** Nature 2014;**506**(7488):328-33.

Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. **Detecting DNA cytosine methylation using nanopore sequencing.** Nature methods 2017;**14**(4):407-10.

Smith ZD, Meissner A. **DNA methylation: roles in mammalian development.** Nature reviews. Genetics 2013;**14**(3):204-20.

Spainhour JC, Lim HS, Yi SV, Qiu P. **Correlation Patterns Between DNA Methylation and Gene Expression in The Cancer Genome Atlas.** Cancer informatics 2019;**18**:117693511982877-1176935119828776.

Spencer DH, Russler-Germain DA, Ketkar S, Helton NM, Lamprecht TL, Fulton RS et al. **CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression.** Cell 2017;**168**(5):801,816.e13.

Spencer DH, Russler-Germain DA, Helton NM, Lamprecht TL, Shinawi M, Westervelt P et al. **DNMT3A-Dependent DNA Methylation May Act As a Tumor Suppressor-Not a Tumor Promoter-during AML Progression.** Blood 2016;**128**(22):1050.

Spencer DH, Young MA, Lamprecht TL, Helton NM, Fulton R, O'Laughlin M et al. **Epigenomic analysis of the HOX gene loci reveals mechanisms that may control canonical expression patterns in AML and normal hematopoietic cells.** Leukemia 2015;**29**(6):1279-89

Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP et al. **Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes.** Blood 2015;**126**(1):9-16.

Stewart SK, Morris TJ, Guilhamon P, Bulstrode H, Bachman M, Balasubramanian S et al. **oxBS-450K: A method for analysing hydroxymethylation using 450K BeadChips.** Methods (San Diego, Calif.) 2015;**72**:9-15.

Strathdee G, Holyoake TL, Alyson Sim A, Parker A, Oscier DG, Melo JV et al. **Inactivation of HOXA Genes by Hypermethylation in Myeloid and Lymphoid Malignancy is Frequent and Associated with Poor Prognosis.** Clinical Cancer Research 2007;**13**(17):5048-55.

Susanto O, Trapani JA, Brasacchio D. **Controversies in granzyme biology.** Tissue antigens 2012;**80**(6):477-87.

Suzuki R, Shimodaira H. **Pvclust: an R package for assessing the uncertainty in hierarchical clustering.** Computer applications in the biosciences 2006;**22**(12):1540-2.

Šálek C. **Akutní myeloidní leukémie.** Postgraduální medicína 2013;**15**(5).

Šestáková Š, Krejčík Z, Folta A, Cerovská E, Šálek C, Merkerová MD et al. **DNA methylation and hydroxymethylation patterns in acute myeloid leukemia patients with mutations in DNMT3A and IDH1/2 and their combinations.** Cancer Biomarkers 2019;**25**(1):43-51.

Šestáková Š, Šálek C, Remešová H. **DNA Methylation Validation Methods: a Coherent Review with Practical Comparison.** Biological Procedures Online 2019;**21**(1).

Taskesen E, Bullinger L, Corbacioglu A, Sanders MA, Erpelinck CAJ, Wouters BJ et al. **Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity.** Blood 2011;**117**(8):2469-75.

The American Cancer Society. **Acute Myeloid Leukemia (AML) Subtypes and Prognostic Factors.** https://www.cancer.org/ 2018(August 21).

Thol F, Bollin R, Gehlhaar M, Walter C, Dugas M, Suchanek KJ et al. **Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications.** Blood 2014;**123**(6):914-20.

Tost J, Gut IG. **DNA methylation analysis by pyrosequencing.** Nature Protocols 2007;**2**(9):2265-75.

Tremblay D, Sokol K, Bhalla S, Rampal R, Mascarenhas JO. **Implications of Mutation Profiling in Myeloid Malignancies-PART 1: Myelodysplastic Syndromes and Acute Myeloid Leukemia.** Oncology (Williston Park, N.Y.) 2018;**32**(4):e38.

Treppendahl MB, Qiu X, Søgaard A, Yang X, Nandrup-Bus C, Hother C et al. **Allelic methylation levels of the noncoding VTRNA2-1 located on chromosome 5q31.1 predict outcome in AML.** Blood 2012;**119**(1):206-16.

Tsai C, Hou H, Tang J, Liu C, Lin C, Chou W et al. **Genetic alterations and their clinical implications in older patients with acute myeloid leukemia.** Leukemia 2016;**30**(7):1485-92.

Tse MY, Ashbury JE, Zwingerman N, King WD, Taylor SA, Pang SC. **A refined, rapid and reproducible high resolution melt (HRM)-based method suitable for quantification of global LINE-1 repetitive element methylation.** BMC research notes 2011;**4**(1):565.

Walter RB, Kantarjian HM, Huang X, Pierce SA, Sun Z, Gundacker HM et al. **Effect of Complete Remission and Responses Less Than Complete Remission on Survival in Acute Myeloid Leukemia: A Combined Eastern Cooperative Oncology Group, Southwest Oncology Group, and M. D. Anderson Cancer Center Study.** Journal of Clinical Oncology 2010;**28**(10):1766-71.

Wang H, Tang K, Liang T, Zhang W, Li J, Wang W et al. **The comparison of clinical and biological characteristics between IDH1 and IDH2 mutations in gliomas.** Journal of experimental & clinical cancer research 2016;**35**(1):86.

Wang H, Wang Y, Qian L, Wang X, Gu H, Dong X et al. **RNF216 contributes to proliferation and migration of colorectal cancer via suppressing BECN1-dependent autophagy.** Oncotarget 2016;**7**(32):51174-83.

Wang X, Gong J, Yu J, Wang F, Zhang X, Yin X et al. **MicroRNA-29a and microRNA-142-3p are regulators of myeloid differentiation and acute myeloid leukemia.** Blood 2012;**119**(21):4992-5004.

Wang Y, Xiao M, Chen X, Chen L, Xu Y, Lv L et al. **WT1 Recruits TET2 to Regulate Its Target Gene Expression and Suppress Leukemia Cell Proliferation.** Molecular cell 2015;**57**(4):662-73.

Ward PS, Patel J, Wise DR, Abdel-Wahab O, Bennett BD, Coller HA et al. **The Common Feature of Leukemia-Associated IDH1 and IDH2 Mutations Is a Neomorphic Enzyme Activity Converting α-Ketoglutarate to 2-Hydroxyglutarate.** Cancer cell 2010;**17**(3):225-34.

Welch J, Ley T, Link D, Miller C, Larson D, Koboldt D et al. **The Origin and Evolution of Mutations in Acute Myeloid Leukemia.** Cell 2012;**150**(2):264-78.

Wertheim GBW, Smith C, Luskin M, Rager A, Figueroa ME, Carroll M et al. **Validation of DNA Methylation to Predict Outcome in Acute Myeloid Leukemia by Use of xMELP.** Clinical chemistry 2015;**61**(1):249-58.

Wertheim GBW, Smith C, Figueroa ME, Kalos M, Bagg A, Carroll M et al. **Microsphere-Based Multiplex Analysis of DNA Methylation in Acute Myeloid Leukemia.** Journal of Molecular Diagnostics, The 2014;**16**(2):207-15.

Wojdacz TK, Borgbo T, Hansen LL. **Primer design versus PCR bias in methylation independent PCR amplifications.** Epigenetics 2009;**4**(4):231-4.

Wojdacz TK, Dobrovic A. **Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation.** Nucleic Acids Research 2007;**35**(6):e41.

Wojdacz TK, Hansen LL, Dobrovic A. **A new approach to primer design for the control of PCR bias in methylation studies.** BMC research notes 2008;**1**(1):54.

Wojdacz TK, Hansen LL. **Reversal of PCR bias for improved sensitivity of the DNA methylation melting curve assay.** BioTechniques 2006;**41**(3):274-8.

Wouters BJ, Delwel R. **Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia.** Blood 2016;**127**(1):42-52.

Wouters B, Löwenberg B, Erpelinck C, Putten W, Valk P, Delwel R. **Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome.** Blood 2009;**113**(13):3088-91.

Zhou J, Lin J, Zhang T, Ma J, Yang L, Wen X et al. **GPX3 methylation in bone marrow predicts adverse prognosis and leukemia transformation in myelodysplastic syndrome.** Cancer Medicine 2017;**6**(1):267-74.

Zhou J, Zhang T, Li X, Ma J, Guo H, Wen X et al. **Methylation-independent CHFR expression is a potential biomarker affecting prognosis in acute myeloid leukemia.** Journal of Cellular Physiology 2018;**233**(6):4707-14.

Zhou J, Zhang T, Wang Y, Yang D, Yang L, Ma J et al. **DLX4 hypermethylation is a prognostically adverse indicator in de novo acute myeloid leukemia.** Tumor Biol 2016;**37**(7):8951-60.

9 SUPPLEMENTARY MATERIAL

Supplementary Table 1 Clinical characteristics of 104 AML patients investigated for the prognostic significance of *GZMB* and *CHFR* methylation

No.	Gender	GZMB methylation			2017 ELN risk stratification	Age at diagnosis [years]	Peripheral leukocytes [10^9/l]	Peripheral blasts [%]	*FLT3*-ITD presence	NPM1 mutation	Transplantation in the 1st complete remission
		Category	[%] Position 1 chr14:25,142,841 (hg38)	[%] Position 2 chr14:25,142,869 (hg38)							
1	Male	HM	69,0	77,1	I/A	55,4	328,8	64	+	-	No
2	Female	IM	33,5	62,4	I	53,5	108,1	62	+	+	No
3	Female	HM	74,8	76,3	I	20,1	15,6	42	-	-	Yes
4	Female	HM	61,1	55,5	I	52,2	59,3	83	+	+	No
5	Male	IM	44,1	69,1	I	39,4	82,3	65	+	+	No
6	Female	HM	59,9	70,8	I	54,1	95,2	64	+	+	No
7	Male	HM	52,3	56,2	I	47,4	53,2	75	+	+	Yes
8	Female	HM	69,1	61,8	F	37,2	144,5	55	+	-	No
9	Female	HM	63,5	65,0	I/A	60,7	7,9	19	-	-	No
10	Female	IM	39,6	50,0	I	49,4	15,3	37	+	+	No
11	Female	HM	52,1	52,0	A	63,9	23,9	56	NT	-	No
12	Male	HM	69,4	71,2	A	64,4	16,4	63	-	-	No
13	Female	HM	79,7	80,5	A	41,8	50,1	73	-	-	No
14	Female	UM	33,9	36,8	I/A	56,6	0,2	20	-	-	Yes
15	Female	UM	32,1	42,5	I	57,7	17,5	12	+	+	No
16	Male	HM	80,8	84,1	F	58,8	161,4	9	-	-	No
17	Female	UM	27,7	28,2	I	53,4	68,9	58	+	-	No
18	Female	UM	16,0	23,4	I	19,2	43,2	63	-	-	No
19	Male	UM	10,6	16,1	F	31,6	8,4	51	-	-	No
20	Male	HM	54,9	72,8	I	53,1	10,8	42	-	-	Yes

21	Female	IM	39,7	76,5	I	50,9	70,4	80	+	+	No
22	Female	HM	47,1	67,6	I	52,6	37,5	40	-	-	No
23	Female	HM	69,8	76,5	F	66,0	6,7	16	-	+	No
24	Male	HM	84,3	94,0	I	45,6	115,3	97	+	-	Yes
25	Male	UM	21,9	35,0	I	65,4	6,1	60	-	-	No
26	Female	HM	64,6	81,3	I	50,3	79,7	73	+	+	No
27	Male	HM	50,4	55,6	I	65,0	8,0	29	-	-	No
28	Male	HM	80,3	91,2	I	37,1	185,1	95	+	-	No
29	Female	UM	14,7	22,4	I	49,8	88,5	84	-	-	No
30	Female	IM	61,5	43,8	F	22,3	54,7	43	-	-	Yes
31	Female	HM	72,6	78,9	F	43,9	9,2	74	-	+	No
32	Male	UM	9,9	16,1	I	36,0	13,0	94	-	-	No
33	Female	HM	83,1	89,5	I	36,2	49,9	99	-	-	No
34	Female	IM	51,2	23,6	I	39,8	45,6	82	-	-	Yes
35	Female	HM	74,1	80,0	I	62,4	4,4	71	-	-	No
36	Female	HM	48,7	57,5	I	30,6	27,7	35	+	-	No
37	Female	HM	65,3	78,4	I	40,3	16,9	42	+	+	No
38	Male	UM	41,5	37,9	I	34,9	3,5	28	-	-	No
39	Male	IM	40,7	47,9	F	55,4	3,3	32	-	+	No
40	Female	HM	63,7	57,9	I	62,5	32,5	29	-	-	No
41	Female	HM	85,5	86,1	I/A	61,9	121,0	90	+	+	No
42	Male	HM	67,8	73,0	A	33,2	5,7	54	-	-	Yes
43	Male	UM	21,3	23,2	F	43,7	54,7	1	-	+	No
44	Male	UM	6,6	10,2	F	32,1	10,3	74	-	-	No
45	Female	HM	66,3	78,2	I	67,4	2,7	65	-	-	No
46	Female	UM	37,5	24,3	I	57,5	50,1	67	-	+	Yes
47	Female	HM	80,1	85,9	F	60,8	67,0	82	-	+	Yes
48	Male	UM	20,9	30,6	I	33,0	49,8	94	-	-	Yes
49	Female	UM	44,8	42,6	I	44,5	78,0	30	+	+	Yes
50	Female	UM	28,4	37,1	I	39,8	33,5	17	+	+	No

51	Male	IM	43,6	46,6	I	51,1	21,7	1	+	+	Yes
52	Male	UM	9,7	11,0	F	64,5	120,3	27	-	+	No
53	Female	HM	60,6	53,6	I/A	62,2	8,4	17	-	-	No
54	Male	HM	83,4	88,9	I/A	24,0	148,5	89	+	+	Yes
55	Male	IM	37,2	47,5	F	35,7	4,9	32	+	-	No
56	Female	UM	10,0	13,0	I	39,9	7,3	38	-	-	No
57	Male	UM	7,0	10,0	A	25,4	37,2	85	+	-	Yes
58	Female	UM	22,6	23,0	I	40,7	29,4	10	+	+	Yes
59	Male	UM	11,3	15,2	A	59,2	2,9	43	-	-	Yes
60	Female	HM	81,9	87,0	F	46,6	30,2	92	-	+	Yes
61	Male	UM	20,9	14,1	F	65,8	3,9	29	-	-	No
62	Female	HM	49,0	53,7	I	54,2	3,6	20	-	-	Yes
63	Male	UM	38,3	32,3	F	29,2	4,7	58	-	-	Yes
64	Female	UM	22,2	20,1	I	32,7	1,7	1	-	-	Yes
65	Female	HM	47,4	64,7	F	46,4	18,3	53	-	+	No
66	Female	HM	70,8	74,6	I	21,0	33,8	44	+	-	Yes
67	Female	IM	48,5	39,2	I	64,3	2,5	6	-	-	No
68	Female	UM	42,5	34,1	I	56,3	2,0	1	-	-	Yes
69	Male	UM	39,3	36,8	I/A	44,5	3,6	1	+	-	No
70	Male	HM	47,4	48,2	I	54,2	120,8	16	-	-	No
71	Male	HM	67,7	72,5	I	52,6	6,3	57	-	-	No
72	Female	IM	52,6	42,1	F	44,6	75,4	29	-	-	No
73	Male	UM	39,0	41,6	F	57,3	65,0	5	-	+	No
74	Female	UM	5,8	6,9	A	62,1	123,3	32	+	-	Yes
75	Male	UM	11,4	8,9	F	20,6	2,20	17	-	-	No
76	Male	IM	46,1	42,2	I	59,6	2,0	9	-	-	No
77	Female	HM	66,1	68,8	F	33,1	29,8	73	-	+	No
78	Female	HM	72,3	79,9	I/A	60,2	140,4	84	+	+	No
79	Male	HM	53,7	57,3	I/A	58,1	31,6	60	-	-	No
80	Male	HM	64,6	68,2	F	57,6	12,4	55	-	+	No

81	Male	UM	22,3	31,8	A	48,5	2,1	3	-	-	Yes
82	Female	UM	20,3	21,1	A	26,2	1,3	10	-	-	Yes
83	Male	UM	13,5	11,6	A	61,7	107,7	85	-	-	No
84	Female	UM	32,2	21,6	A	60,5	1,4	1	-	-	No
85	Male	HM	79,4	84,6	I	38,5	10,0	91	-	-	Yes
86	Male	HM	79,7	86,5	I	63,4	25,6	92	-	-	No
87	Female	HM	54,0	60,2	I	57,2	271,4	92	+	-	Yes
88	Male	UM	14,2	6,7	F	55,3	11,3	73	-	-	No
89	Male	HM	69,1	76,6	I	62,7	8,1	52	-	-	No
90	Male	IM	40,7	46,4	I	61,4	2,9	5	+	-	No
91	Male	UM	27,5	23,2	I	63,4	1,5	2	-	-	No
92	Male	IM	57,8	29,5	A	56,4	8,0	86	-	-	Yes
93	Male	UM	40,8	43,7	I	56,1	1,2	0	-	-	Yes
94	Male	UM	8,2	10,4	F	33,5	28,7	72	-	-	No
95	Male	HM	63,8	52,9	A	58,5	1,3	8	-	-	Yes
96	Female	UM	41,0	34,0	I	47,9	2,7	1	-	-	No
97	Male	HM	70,0	73,5	I	61,9	2,1	66	-	-	Yes
98	Male	IM	47,9	39,0	A	55,4	1,6	14	-	-	No
99	Female	UM	15,9	17,7	I	21,0	4,1	11	+	-	Yes
100	Male	HM	63,9	47,0	F	40,9	55,2	89	-	+	Yes
101	Male	UM	25,0	26,0	I	30,0	6,6	85	+	-	Yes
102	Male	UM	23,4	24,5	I	59,9	1,2	20	-	-	No
103	Female	HM	86,9	91,5	I	65,8	154,4	99	-	-	Yes
104	Female	HM	54,6	53,6	A	38,5	10,0	0	-	NT	No

IM - intermediately methylated, UM – unmethylated, HM – hypermethylated, A – Adverse, I – Intermediate, F – Favorable, NT - not tested

Supplementary Table 2 Clinical characteristics of 128 AML patients from the test cohort investigated with the DNA methylation sequencing panel

ID	Age [years]	Sex (Male/Female)	AML classification FAB	Leukocytes [10^9/l]	Blasts in peripheral blood [%]	Blasts in bone marrow [%]	Cytogenetics (Grimwade 2010)	Transplantation in the 1st CR	Transplantation time [months]	Relapse	CR in the 1st induction	Number of mutations	*FLT3*-ITD	DNMT3A^{mut}	*IDH1/2*^{mut}	*TET2*^{mut}	*ASXL1*^{mut}	*NRAS*^{mut}	*TP53*^{mut}	*NPM1*^{mut}	*CEBPA*^{mut}	*RUNX1*^{mut}	OS [months]	Death	EFS [months]	Relapse/Death	MethScore	MethScore group
A1	57	M	4	44	0.0	17.8	2	1	6.9	0	1	3	0	1	0	0	0	0	0	1	0	0	48.6	0	43.3	0	669.6	low
A2	64	F	2	64	6.3	15.0	2	0	0.0	1	0	0	0	0	0	NA	NA	NA	0	0	0	NA	35.7	1	34.0	1	1013.7	high
A3	56	F	8	55	15.4	38.0	2	0	0.0	0	1	4	1	1	1	0	0	0	0	1	0	0	83.2	0	82.0	0	606.7	low
A4	58	M	8	81	2.0	24.2	2	0	0.0	0	0	1	0	0	0	0	0	0	0	0	0	0	8.0	1	4.0	1	1370.6	high
A5	33	F	5	23	0.5	67.0	2	1	5.8	0	1	1	0	0	0	1	0	0	0	0	0	0	8.2	1	7.3	1	1180.9	high
A6	45	M	5	83	100.0	0.0	2	0	0.0	1	1	1	1	0	0	0	0	0	0	0	0	0	9.9	1	2.2	1	993.2	high
A7	36	M	8	52	23.0	20.8	2	0	0.0	0	1	4	1	1	0	0	0	0	0	1	0	0	75.1	0	74.2	0	126.1	low
A8	40	M	8	9	0.0	30.8	3	0	0.0	0	1	1	0	0	0	0	0	0	1	0	0	0	30.3	1	1.5	1	1203.9	high
A9	49	M	8	58	2.5	24.0	3	1	1.6	1	0	3	0	1	0	1	0	0	0	0	0	0	26.5	1	22.7	1	1002.4	high
A10	58	M	4	39	11.0	36.0	2	0	0.0	1	1	4	0	1	1	0	0	1	0	1	0	0	11.5	1	8.6	1	1114.6	high
A11	62	M	5	136	14.5	39.2	2	0	0.0	0	1	3	0	1	0	0	0	0	0	1	0	0	25.0	1	24.0	1	1088.1	high
A12	63	F	5	132	0.0	77.0	2	0	0.0	0	NA	3	0	0	0	0	1	0	0	0	0	0	1.1	1	0.0	1	1411.0	high
A13	53	F	4	1	63.5	78.8	2	0	0.0	0	0	3	1	1	0	1	0	0	0	0	0	0	1.3	1	0.0	1	1365.8	high
A14	52	F	1	92	98.0	96.4	2	0	0.0	1	0	4	1	0	0	0	1	0	0	1	0	1	12.1	1	0.0	1	1191.7	high
A15	36	M	5	11	15.2	93.2	3	0	0.0	0	0	0	0	0	0	NA	NA	NA	0	0	0	NA	0.7	1	0.0	1	859.2	high
A16	52	M	1	7	3.4	68.2	2	1	5.3	0	0	1	0	1	0	0	0	0	0	0	0	0	59.5	0	57.2	0	542.2	low
A17	56	M	6	59	0.0	1.2	2	0	0.0	0	0	0	0	0	0	NA	NA	NA	NA	0	0	NA	3.9	1	0.3	1	1427.6	high
A18	58	M	6	25	1.4	9.8	2	1	3.9	0	1	0	0	0	0	NA	NA	NA	NA	0	0	NA	60.1	0	58.7	0	658.7	low
A19	68	F	5	66	1.0	23.8	2	0	0.0	1	1	3	1	1	0	0	0	0	0	1	0	0	54.0	0	29.7	1	490.7	low

A20	42	F	2	12	4.0	49.0	3	1	4.4	0	0	0	0	NA	NA	NA	NA	NA	NA	0	0	NA	51.6	0	49.7	0	667.1	low
A21	57	M	2	49	64.5	67.4	2	1	6.0	0	0	1	0	1	0	0	0	0	0	0	0	0	49.2	0	47.5	0	572.3	low
A22	65	F	6	6	2.5	5.4	2	1	5.1	0	1	1	0	NA	NA	NA	NA	NA	0	0	0	NA	48.6	0	47.5	0	582.7	low
A23	30	F	4	51	27.5	72.4	2	1	5.5	1	1	4	0	1	0	0	1	1	0	1	0	0	65.9	0	16.3	1	638.9	low
A24	27	M	1	18	2.0	22.0	3	1	5.9	0	0	0	0	0	0	0	0	0	0	0	0	0	44.4	0	42.2	0	723.3	low
A25	61	F	7	24	0.0	67.2	3	0	0.0	0	1	3	0	1	0	0	0	0	1	0	1	0	8.5	1	6.4	1	1225.4	high
A26	53	M	2	NA	27.0	73.0	2	1	0.3	0	0	1	0	1	NA	NA	NA	NA	NA	NA	NA	NA	4.0	1	0.0	1	1353.9	high
A27	62	M	8	96	9.9	23.4	3	1	5.3	1	0	0	0	0	0	0	0	0	0	0	0	0	10.6	1	5.7	1	1144.3	high
A28	55	F	8	94	3.2	NA	NA	1	3.9	1	0	0	0	NA	NA	NA	NA	NA	0	NA	NA	NA	8.3	1	0.0	1	1125.1	high
A29	48	F	5	87	12.0	59.6	2	1	5.1	1	1	4	1	1	0	0	0	1	0	1	0	0	9.9	1	7.4	1	1016.7	high
A30	62	M	8	8	0.5	17.8	2	1	3.9	0	1	0	0	0	0	NA	NA	NA	NA	0	0	NA	36.9	0	34.2	0	496.0	low
A31	61	M	4	29	9.5	26.8	2	0	0.0	0	0	3	1	0	0	0	1	0	0	0	0	1	6.7	1	0.0	1	1478.6	high
A32	51	F	4	119	43.0	47.4	2	1	6.6	1	1	3	1	1	0	0	0	0	0	1	0	0	55.8	0	20.4	1	1088.6	high
A33	43	M	4	88	40.6	46.4	2	1	16.7	0	1	2	0	0	1	NA	NA	NA	NA	1	NA	NA	63.5	0	46.8	0	939.6	high
A34	46	F	2	53	46.0	46.0	2	0	0.0	1	1	2	0	1	0	0	0	0	0	1	0	0	86.4	0	60.3	1	739.9	low
A35	62	M	5	115	0.5	48.6	3	0	0.0	1	1	4	0	1	0	0	0	1	0	1	0	0	3.8	1	2.5	1	950.4	high
A36	34	F	4	126	5.0	21.6	2	0	0.0	1	1	3	0	1	0	0	0	0	0	1	0	0	21.7	1	8.7	1	1255.4	high
A37	64	M	4	133	77.0	80.6	2	0	0.0	0	0	4	1	1	1	0	0	0	0	1	0	0	1.8	1	0.0	1	1385.6	high
A38	58	M	1	85	60.4	92.8	2	0	0.0	0	0	2	0	1	1	0	0	0	0	0	0	0	4.1	1	0.0	1	1563.1	high
A39	48	F	5	106	4.5	53.2	2	1	7.5	1	1	2	0	1	0	0	0	0	0	1	0	0	43.9	0	22.9	1	613.4	low
A40	38	F	8	15	8.8	23.6	2	0	0.0	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	0	NA	40.6	0	31.8	1	597.1	low
A41	62	F	2	17	19.1	52.8	3	1	4.6	0	1	1	0	0	0	NA	NA	NA	NA	0	0	NA	9.2	1	8.1	1	1126.5	high
A42	43	F	4	135	9.5	35.0	2	1	4.8	0	0	2	1	NA	NA	NA	NA	NA	NA	1	0	NA	38.0	0	35.8	0	887.6	high
A43	22	F	5	95	0.9	93.2	3	0	0.0	1	1	0	0	NA	NA	NA	NA	NA	NA	0	0	NA	5.9	1	0.5	1	1261.9	high
A44	57	F	5	77	92.0	92.0	2	1	5.0	0	0	3	1	1	0	0	0	0	0	0	0	0	68.7	0	66.1	0	1088.0	high
A45	26	F	1	14	9.5	81.8	3	1	6.0	0	0	2	0	1	1	0	0	0	0	0	0	0	73.4	0	72.1	0	535.7	low
A46	59	M	6	13	8.3	12.6	3	1	5.1	0	0	0	0	NA	NA	NA	NA	NA	0	0	NA	NA	67.5	0	64.3	0	999.8	high
A47	57	F	5	68	36.0	95.8	2	1	5.4	0	1	0	0	0	0	0	0	0	0	0	0	0	16.3	1	14.9	1	775.9	low
A48	27	M	4	76	2.0	19.6	2	1	3.5	0	1	1	0	NA	NA	NA	NA	NA	NA	0	0	NA	3.9	1	3.0	1	758.7	low
A49	40	F	8	31	22.4	22.4	2	1	3.5	0	1	2	0	1	0	NA	NA	NA	NA	1	0	NA	75.5	0	74.1	0	395.0	low

A50	54	M	4	34	19.0	19.4	2	0	0.0	1	1	0	0	NA	NA	NA	NA	NA	NA	0	0	NA	9.6	1	4.2	1	1134.2	high
A51	61	F	2	16	0.5	70.2	3	0	0.0	0	0	1	0	0	0	0	0	0	1	0	0	0	16.2	1	0.0	1	1168.7	high
A52	60	F	8	50	25.0	NA	2	1	2.3	0	1	5	1	1	0	0	0	1	0	1	0	0	10.7	1	9.5	1	903.6	high
A53	61	M	2	67	5.0	22.6	2	0	0.0	0	0	4	1	1	1	0	0	0	0	0	0	0	4.3	1	0.0	1	1062.1	high
A54	63	M	1	20	1.9	68.8	2	0	0.0	1	1	2	0	0	1	0	0	0	0	0	0	0	48.4	1	8.4	1	768.8	low
A55	58	M	8	57	2.9	26.6	3	0	0.0	0	0	1	0	0	0	0	0	0	1	0	0	0	19.1	1	0.0	1	852.8	high
A56	34	M	2	78	72.0	44.8	1	0	0.0	1	1	0	0	0	0	0	0	0	0	0	0	0	32.1	1	1.3	1	908.2	high
A57	21	F	1	93	11.0	53.6	2	1	5.3	0	1	1	1	0	0	0	0	0	0	0	0	0	64.9	0	63.8	0	385.9	low
A58	41	M	1	108	89.0	85.2	2	1	4.1	0	0	3	0	0	0	1	0	1	0	1	0	0	63.7	0	61.6	0	857.9	high
A59	30	M	2	112	84.5	45.0	2	1	4.5	1	1	2	1	0	1	0	0	0	0	0	0	0	62.8	0	13.1	1	549.4	low
A60	60	M	1	10	20.0	69.8	2	0	0.0	1	1	3	0	1	1	0	0	0	0	0	0	0	51.8	1	11.1	1	569.7	low
A61	39	F	4	134	0.0	20.2	3	0	0.0	0	0	1	0	0	0	0	0	0	1	0	0	0	2.1	1	0.0	1	1231.2	high
A62	53	F	4	110	62.0	46.4	2	0	0.0	0	1	4	0	1	1	0	0	1	0	1	0	0	16.9	1	15.8	1	1060.7	high
A63	69	F	1	37	96.0	93.2	2	0	0.0	0	0	4	1	0	0	1	0	0	0	1	0	0	4.2	1	0.0	1	1327.0	high
A64	60	M	1	35	94.0	96.4	2	1	3.9	1	0	6	1	1	1	1	0	0	0	1	1	0	22.8	1	16.8	1	1125.2	high
A65	60	M	2	103	64.0	69.8	2	0	0.0	0	0	3	0	0	1	0	0	0	0	1	0	0	0.9	1	0.0	1	1227.2	high
A66	51	F	5	5	2.9	88.8	NA	1	3.4	1	1	1	0	0	0	0	0	0	0	0	0	0	37.4	1	27.7	1	610.9	low
A67	61	F	1	97	86.0	79.6	2	0	0.0	1	1	3	0	1	1	0	0	0	0	0	1	0	24.1	1	6.3	1	745.8	low
A68	62	M	1	19	3.0	49.2	2	1	4.4	0	0	0	0	0	0	0	0	0	0	0	0	0	52.0	0	49.8	0	245.8	low
A69	62	F	4	98	1.5	21.0	3	1	6.2	1	0	1	0	NA	NA	NA	NA	NA	1	0	0	NA	12.5	1	7.1	1	852.3	high
A70	59	F	4	99	73.5	71.8	3	0	0.0	0	0	3	0	0	0	1	1	1	0	0	0	0	6.1	1	0.0	1	1415.4	high
A71	28	M	1	122	0.0	69.6	2	1	6.8	0	1	2	1	0	0	0	0	0	0	0	0	0	61.7	0	60.6	0	326.8	low
A72	67	M	4	120	13.0	24.4	3	0	0.0	0	0	3	0	0	0	0	0	0	1	0	0	0	0.8	1	0.0	1	1185.9	high
A73	39	F	4	69	7.5	25.4	3	1	3.4	1	1	2	0	0	0	0	0	0	0	0	1	0	61.0	0	5.3	1	717.4	low
A74	62	M	5	4	0.5	82.2	2	0	0.0	0	1	0	0	0	0	0	0	0	0	0	0	0	3.5	1	2.6	1	993.6	high
A75	63	M	8	28	19.0	NA	3	0	0.0	0	0	1	0	NA	NA	NA	NA	NA	1	NA	NA	NA	5.3	1	0.0	1	1187.5	high
A76	43	M	4	101	20.5	42.0	3	0	0.0	1	0	3	0	0	0	0	0	0	0	0	1	1	26.4	1	1.8	1	1187.3	high
A77	61	F	1	3	0.0	80.8	2	1	3.5	0	1	3	1	0	1	0	0	0	0	1	0	0	49.0	0	48.1	0	98.4	low
A78	58	M	1	71	NA	94.0	2	1	7.0	0	0	3	0	0	0	1	0	1	0	1	0	0	52.6	0	51.0	0	817.2	high
A79	66	F	5	63	0.4	88.6	2	0	0.0	1	1	1	0	NA	0	NA	NA	NA	NA	1	0	NA	5.7	1	1.8	1	874.2	high

A80	28	F	4	121	14.5	35.2	2	0	0.0	0	0	4	1	1	0	1	0	0	0	1	0	0	3.2	1	0.0	1	1289.3	high
A81	38	F	1	91	84.0	90.6	1	1	5.2	1	1	1	0	0	0	0	0	0	0	0	0	0	42.3	0	1.1	1	588.5	low
A82	65	F	5	117	22.5	88.0	2	0	0.0	0	1	4	0	1	0	1	0	0	0	1	0	0	41.1	0	39.5	0	656.0	low
A83	58	M	2	32	12.5	41.6	2	1	4.9	0	0	1	0	1	0	NA	NA	NA	NA	0	0	NA	36.6	1	34.4	1	642.9	low
A84	57	M	5	118	4.5	34.2	2	0	0.0	0	1	2	0	0	0	0	0	0	0	1	0	0	75.8	0	74.9	0	504.0	low
A85	63	F	1	116	98.5	97.8	2	0	0.0	0	0	4	1	0	1	0	0	0	0	0	0	1	0.8	1	0.0	1	1303.9	high
A86	52	M	8	40	7.0	24.8	2	0	0.0	0	0	1	0	0	0	0	0	0	0	0	0	1	43.1	0	39.8	0	424.1	low
A87	21	F	2	86	43.0	51.0	2	1	5.4	1	0	2	1	0	0	0	0	0	0	0	0	0	82.9	0	1.4	1	611.7	low
A88	63	M	1	104	67.0	60.8	2	0	0.0	0	0	2	0	0	0	0	1	1	0	0	0	0	7.6	1	0.0	1	1368.5	high
A89	37	M	4	79	36.0	55.4	1	1	17.9	0	1	2	1	0	0	0	0	0	0	0	0	0	79.6	0	78.5	0	396.5	low
A90	62	F	4	36	32.0	69.0	3	1	4.8	0	1	2	1	0	0	0	0	0	0	1	0	0	10.2	1	9.3	1	824.1	high
A91	53	M	2	111	57.0	54.6	2	0	0.0	1	1	4	1	0	1	0	0	1	0	0	0	0	41.4	1	6.8	1	538.5	low
A92	31	M	2	75	17.4	26.0	1	0	0.0	0	1	0	0	0	0	0	0	0	0	0	0	0	76.4	0	75.0	0	281.6	low
A93	60	M	4	26	8.6	48.6	2	0	0.0	1	1	1	0	0	0	0	0	0	0	0	0	0	19.7	1	3.0	1	434.0	low
A94	33	F	2	80	73.0	69.2	2	0	0.0	0	1	2	0	0	0	0	0	1	0	1	0	0	35.8	0	34.8	0	503.2	low
A95	42	M	8	61	1.0	20.2	3	0	0.0	1	NA	1	0	NA	NA	NA	NA	NA	0	0	0	NA	75.4	1	65.3	1	857.7	high
A96	58	M	2	33	55.0	75.0	2	0	0.0	1	1	3	0	0	1	0	0	0	0	1	0	1	76.3	0	38.0	1	428.5	low
A97	54	M	8	123	1.0	21.6	2	1	6.8	0	1	1	0	0	0	0	0	0	0	0	0	0	66.5	0	65.3	0	409.9	low
A98	55	M	2	30	73.0	53.8	1	0	0.0	1	1	0	0	0	0	0	0	0	0	0	0	0	8.3	1	3.0	1	800.7	low
A99	63	M	2	130	52.0	31.6	2	0	0.0	0	0	1	0	0	0	1	0	0	0	0	0	0	60.6	0	57.8	0	278.7	low
A100	39	M	2	27	91.0	64.6	2	1	4.5	0	0	1	0	1	0	0	0	0	0	0	0	0	74.5	0	73.1	0	326.9	low
A101	56	M	1	124	86.0	95.0	3	1	3.3	1	1	4	0	1	1	0	0	0	1	0	0	1	10.0	1	8.2	1	1022.4	high
A102	62	M	1	60	65.6	79.4	2	1	4.1	0	0	2	0	0	1	0	0	0	0	0	1	0	65.9	0	64.5	0	289.5	low
A103	63	M	1	74	92.0	68.4	2	0	0.0	0	0	2	0	0	0	0	0	0	1	0	1	0	9.4	1	0.0	1	993.3	high
A104	36	F	6	22	10.0	4.8	2	0	0.0	0	1	1	0	0	0	NA	NA	NA	NA	1	0	NA	60.5	0	59.4	0	288.6	low
A105	55	F	2	109	75.0	74.4	2	1	6.7	0	1	2	1	0	0	0	0	0	0	1	0	0	58.9	0	58.0	0	661.6	low
A106	24	F	1	48	89.5	93.6	2	1	4.6	0	1	4	1	0	1	0	1	0	0	1	0	0	46.7	0	45.8	0	744.2	low
A107	37	M	1	114	83.0	90.2	3	0	0.0	1	0	1	0	0	0	0	0	1	0	0	0	0	15.7	1	12.4	1	596.4	low
A108	39	M	4	72	61.6	61.6	2	1	5.6	0	1	2	1	0	0	0	0	0	0	0	0	1	53.6	0	52.5	0	429.0	low
A109	43	M	2	38	83.5	85.0	2	0	0.0	0	1	1	0	0	0	0	0	0	0	1	0	0	52.5	0	51.3	0	619.6	low

A110	50	M	1	62	53.4	94.0	3	1	4.9	0	0	3	0	0	0	0	0	0	0	1	0	1	48.7	0	46.4	0	307.3	low
A111	52	F	1	82	75.1	75.2	2	0	0.0	0	1	3	0	1	1	0	0	0	0	1	0	0	47.4	0	46.5	0	841.7	high
A112	61	F	4	70	49.0	47.2	3	1	4.4	1	0	2	0	0	0	0	0	1	1	0	0	0	7.4	1	4.3	1	1151.4	high
A113	66	F	1	45	98.5	92.6	2	1	6.6	0	0	1	0	0	1	0	0	0	0	0	0	0	58.8	0	56.4	0	566.3	low
A114	28	F	1	102	0.0	88.0	2	0	0.0	1	1	1	0	0	0	0	0	0	0	0	1	0	62.1	0	16.9	1	345.6	low
A115	55	F	4	100	25.5	32.2	2	0	0.0	1	1	7	0	1	1	0	0	1	0	1	0	0	35.3	1	18.2	1	927.2	high
A116	37	F	4	65	1.4	31.4	3	0	0.0	0	0	0	0	NA	NA	NA	NA	NA	NA	0	0	NA	2.4	1	0.0	1	716.9	low
A117	56	F	4	128	46.0	56.8	2	0	0.0	0	0	2	1	0	0	0	0	0	0	0	0	0	4.4	1	0.0	1	1127.9	high
A118	45	F	4	127	29.0	37.8	1	0	0.0	0	1	2	0	0	0	0	1	1	0	0	0	0	80.0	0	78.9	0	484.1	low
A119	35	F	4	46	44.0	58.2	1	1	21.4	0	1	1	0	0	0	0	0	0	0	0	0	0	47.0	0	45.7	0	180.5	low
A120	21	M	1	56	17.0	59.6	1	0	0.0	0	1	1	0	0	0	0	1	0	0	0	0	0	78.0	0	76.6	0	72.1	low
A121	37	F	2	131	60.0	67.6	2	0	0.0	0	1	5	0	1	0	0	0	1	0	1	0	0	40.8	0	39.9	0	574.9	low
A122	64	F	2	NA	NA	NA	2	0	0.0	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	42.5	1	0.0	1	685.9	low
A123	61	M	1	73	89.5	95.6	2	1	3.9	0	1	6	1	1	1	0	0	0	0	1	0	0	12.6	1	11.6	1	1031.8	high
A124	41	M	2	84	23.2	23.2	1	1	26.7	0	1	0	0	0	0	0	0	0	0	0	0	0	45.6	0	44.2	0	400.8	low
A125	60	F	1	41	83.5	81.2	2	0	0.0	0	0	4	1	1	1	0	0	0	0	1	0	0	5.3	1	0.0	1	1184.5	high
A126	56	M	6	11	0.0	32.0	2	1	5.1	0	1	1	0	0	0	0	0	0	0	0	0	0	70.0	0	69.0	0	220.6	low
A127	55	M	2	21	13.6	30.6	3	0	0.0	1	0	2	0	0	0	0	0	0	1	0	0	0	14.3	1	4.1	1	745.6	low
A128	61	M	8	2	2.6	40.2	3	0	0.0	0	0	0	0	0	0	0	0	0	0	0	0	0	31.7	1	0.0	1	923.3	high

M - male, F - female, CR - complete remission, FAB – French American British classification

Supplementary Table 3 Clinical characteristics of 50 AML patients from the validation cohort investigated with the DNA methylation sequencing panel

ID	Age [years]	Sex (Male/Female)	AML classification FAB	Leukocytes [10^9/l]	Blasts in peripheral blood [%]	Blasts in bone marrow [%]	Cytogenetics (Grimwade 2010)	Transplantation in the 1st CR	Transplantation time [months]	Relapse	CR in the 1st induction	Number of mutations	*FLT3*-ITD	DNMT3A^{mut}	*IDH1/2*^{mut}	*TET2*^{mut}	*ASXL1*^{mut}	*NRAS*^{mut}	*TP53*^{mut}	*NPM1*^{mut}	*CEBPA*^{mut}	*RUNX1*^{mut}	OS [months]	Death	EFS [months]	Relapse/Death	MethScore	MethScore group
B48	66	F	M8	75.5	2	55.8	2	0	0	0	0	3	1	0	0	0	1	1	0	0	0	0	2.1	1	0.0	1	1413.9	high
B07	33	M	M4	14.7	7	51.2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.6	1	0.0	1	897.1	high
B03	75	F	M1	1.64	8	29.8	3	0	0	0	0	2	0	0	0	0	0	0	1	0	0	0	3.4	1	0.0	1	606.9	low
B41	46	F	M8	25.9	50	36.4	2	0	0	0	0	3	1	1	0	0	0	0	0	1	0	0	3.8	1	0.0	1	967.2	high
B43	69	M	M8	12.1	27	20.4	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3.9	1	0.0	1	552.3	low
B32	70	M	M4	22.6	27	50.4	2	0	0	1	1	5	0	1	1	0	0	0	0	1	0	0	4.9	1	0.7	1	593.2	low
B44	58	F	M2	9.36	42.5	61.8	2	0	0	1	1	4	0	1	0	1	0	0	0	1	1	0	6.1	1	4.2	1	1088.7	high
B17	47	M	M2	14.4	84.5	65.8	3	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	6.9	1	0.0	1	745.9	high
B25	73	F	M8	0.69	15.5	46	2	0	0	1	0	2	0	1	1	0	0	0	0	0	0	0	7.0	1	3.0	1	629.6	low
B11	57	F	M4	106	52	69.6	2	0	0	0	0	2	1	1	0	0	0	0	0	0	0	0	7.0	1	0.0	1	708.3	high
B37	67	F	M8	23.1	17	24.4	2	0	0	1	1	4	1	1	0	0	0	0	0	1	0	0	8.5	1	4.1	1	974.9	high
B24	65	M	M6	36.7	29.5	35.8	3	1	3.6	1	0	1	0	0	0	0	0	0	1	0	0	0	8.9	1	5.4	1	669.9	low
B35	53	M	M5	29	9.5	82.6	2	0	0	1	1	3	1	1	0	0	0	0	0	1	0	0	10.0	1	4.6	1	1007.2	high
B30	67	M	M8	7.84	10.5	20	2	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0	10.2	1	5.6	1	421.1	low
B06	32	M	M1	58.4	91.5	82.2	3	1	4.0	1	0	1	0	0	0	0	0	1	0	0	0	0	10.7	1	6.4	1	666.2	low
B12	56	F	M2	9.06	11	41.4	3	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	11.3	1	0.0	1	675.3	high
B02	59	F	M4	41	27	59	2	0	0	1	1	5	1	1	0	0	0	1	0	1	0	0	12.1	1	8.7	1	745.3	high
B01	63	M	M4	7.99	15.5	27.8	2	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	12.6	1	6.5	1	667.0	low
B40	66	F	M4	19.4	24.5	25	3	1	3.9	1	0	1	0	0	0	0	1	0	0	0	0	0	14.1	1	7.8	1	814.7	high
B29	45	F	M8	1.82	1.5	57.2	2	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	14.3	1	8.1	1	700.4	high
B34	69	M	M8	121	4.5	33.4	2	0	0	1	1	6	0	0	0	0	1	1	0	0	0	1	14.4	1	5.5	1	1163.6	high

B10	40	F	M2	54.7	77.5	54	2	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	14.6	1	7.1	1	1003.7	high
B45	47	M	M4	83.4	19.5	39.8	2	1	7.6	1	1	2	1	0	0	0	0	0	0	1	0	0	17.2	1	8.5	1	721.9	high
B14	68	M	M8	12.7	7	28.2	2	0	0	1	1	2	0	0	0	1	0	0	0	1	0	0	21.7	0	7.9	1	507.8	low
B38	54	F	M8	105	7.5	34.4	2	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	29.3	1	7.5	1	631.0	low
B09	43	M	M2	39.6	53.5	74.8	2	1	3.9	1	0	1	0	0	0	0	0	0	0	0	0	0	31.8	1	24.9	1	390.1	low
B46	65	M	M8	2.47	1.5	25	2	0	0	1	1	2	0	0	0	0	1	0	0	0	0	0	33.0	1	16.8	1	699.3	high
B39	59	F	M2	182	85	63.4	2	0	0	1	1	3	0	0	1	0	0	0	0	1	0	0	33.2	1	7.4	1	895.1	high
B50	24	F	M8	64.5	65	28.8	3	0	0	1	1	4	1	0	1	0	0	1	0	1	0	0	33.8	1	5.8	1	670.0	high
B20	25	M	M4	96.1	35.5	27.8	2	1	4.6	0	1	1	0	0	0	0	0	0	0	1	0	0	38.7	0	37.7	0	438.8	low
B04	70	M	M2	30.3	68.5	90	2	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	48.4	1	46.9	1	725.0	high
B49	50	M	M2	1.28	11.5	45	3	1	5.1	0	0	2	0	0	0	0	0	1	0	0	0	0	51.8	0	48.9	0	244.4	low
B47	60	F	M8	1.93	6	25	2	1	4.1	0	0	1	0	0	0	0	0	0	0	0	0	0	54.5	0	52.7	0	770.3	high
B42	54	F	M8	218	57.5	NA	2	1	3.6	0	1	3	1	1	0	0	0	0	0	1	0	0	55.8	0	54.5	0	1194.7	high
B36	62	M	M1	110	97	82	2	1	4.9	0	1	0	0	0	0	0	0	0	0	0	0	0	58.4	0	57.2	0	505.4	low
B33	60	F	M2	19.9	18	30.4	1	0	0	0	1	2	0	1	0	0	0	0	0	0	0	0	59.4	0	58.2	0	422.6	low
B26	70	F	M1	147	84.5	91.11	2	0	0	0	1	3	1	0	0	1	0	0	0	1	0	0	61.1	0	59.9	0	946.8	high
B31	62	M	M6	1.24	6	74	2	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	61.5	0	59.9	0	506.6	low
B22	64	F	M1	3.95	69	89.4	2	1	4.4	0	1	5	1	0	1	1	0	0	0	1	0	0	62.2	0	61.2	0	825.4	high
B08	67	F	M2	1.35	0.5	25.2	2	0	0	0	1	2	0	0	0	0	0	1	0	1	0	0	62.9	0	61.9	0	629.4	low
B28	54	F	M1	133	94.5	91.8	2	1	4.8	0	1	3	1	0	1	0	0	0	0	1	0	0	63.1	0	61.9	0	605.3	low
B18	47	F	M8	0.81	28	50	2	1	5.6	0	0	1	0	0	0	0	0	0	0	0	0	1	63.7	0	61.8	0	508.6	low
B27	47	M	M1	218	92	86.4	2	1	4.7	0	1	3	0	0	0	0	0	1	0	0	1	0	63.8	0	62.6	0	426.5	low
B23	28	M	M5	16.1	49.5	77.4	3	1	4.6	0	1	1	0	0	0	0	0	0	0	0	0	0	64.7	0	63.5	0	731.9	high
B21	31	F	M4	3.93	16.5	40	1	0	0	1	1	2	0	0	0	0	0	1	0	0	0	0	64.9	0	10.2	1	768.5	high
B19	72	F	M4	44.9	17	37	1	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	65.2	0	64.2	0	465.9	low
B15	64	M	M5	185	21.5	72.2	2	1	5.6	0	1	2	1	0	0	0	0	0	0	1	0	0	65.9	0	64.8	0	661.9	low
B16	59	F	M1	2.17	8	71.4	2	1	6.5	0	1	2	0	0	1	0	0	0	0	1	0	0	67.6	0	66.3	0	399.6	low
B13	66	M	M8	34.9	87	73.4	2	0	0	1	1	2	0	0	0	0	0	0	0	1	0	0	69.0	0	11.1	1	655.9	low
B05	29	M	M8	2.53	15	26.4	2	1	3.2	0	1	0	0	0	0	0	0	0	0	0	0	0	74.4	0	73.3	0	637.5	low

M - male, F - female, CR - complete remission, FAB – French American British classification

Supplementary Table 4 Genomic regions targeted by the methylation sequencing panel (according to Human GRCh37/hg19 assembly)

Chromosome	Start	End	Gene annotation	Size (bp)	Source
chr1	208083683	208085683	*CD34*	2001	
chr1	113249025	113251025	*RHOC*	2001	
chr7	30028425	30030425	*SCRN1*	2001	
chr5	76113832	76115832	*F2RL1*	2001	Marcucci *et al.*, 2014
chr8	94711772	94713772	*FAM92A1*	2001	
chr21	26933456	26935456	*MIR155HG*	2001	
chr13	42534221	42536221	*VWA8*	2001	
chr11	118763110	118763426	*BLR1 (CXCR5)*	316	
chr17	2208021	2208391	*SMG6/SRR*	370	
chr19	37958559	37958860	*ZNF569*	301	
chr20	814970	815202	*FAM110A*	232	
chr13	53028642	53029495	*CKAP*	853	
chr20	11898555	11898849	*BTBD3*	294	
chr3	129274773	129275235	*PLXND1*	462	
chr14	24867489	24867729	*KIAA1305*	240	
chr15	50838542	50839225	*USP50*	683	
chr6	34856156	34857019	*ANKS1A*	863	Figueroa *et al.*, 2010;
chr20	11898849	11899205	*BTBD3*	356	Wertheim et al., 2014;
chr18	5293969	5294770	*ZFP161*	801	Luskin et al., 2016
chr2	158114266	158115184	*GALNT5*	918	
chr16	68345197	68345691	*SLC7A6OS/PRMT7*	494	
chr20	11899205	11899843	*BTBD3*	638	
chr20	32274469	32275009	*E2F1*	540	
chr14	106354882	106355276	*FAM30A*	394	
chrX	48795887	48797005	*OUTD5*	1118	
chr1	32739167	32739750	*LCK*	583	
chr12	6233715	6234255	*VWF*	540	
chr3	8542436	8543339	*LMCD1*	903	
chr14	100069833	100070055	*CCDC85C*	223	
chr3	239867	240082	*CHL1*	216	
chr9	23820756	23820974	*ELAVL2*	219	
chr7	143582381	143582625	*FAM115A*	245	
chr10	128995052	128995292	*FAM196A*	241	
chr7	1096275	1096523	*GPR146*	249	
chr6	110300208	110300421	*GPR6*	214	Li *et al.*, 2016
chr20	62198506	62198740	*HELZ2*	235	
chr6	19837914	19838139	*ID4*	226	
chr6	19837413	19837624	*ID4*	212	
chr10	6104037	6104302	*IL2RA*	266	
chr2	42720178	42720398	*KCNG3*	221	
chr20	26189963	26190193	*LOC284801*	231	

chr15	89921302	89921553	*LOC254559*	252	
chr15	89921743	89922005	*LOC254559*	263	
chr2	101437038	101437287	*NPAS2*	250	
chr5	140346432	140346663	*PCDHAC2*	232	
chr5	140346119	140346352	*PCDHAC2*	234	
chr5	138729890	138730214	*PROB1*	325	
chr17	11145296	11145505	*SHISA6*	210	
chr10	50818800	50819045	*SLC18A3*	246	
chr12	93967113	93967390	*SOCS2*	278	
chr1	231298882	231299100	*TRIM67*	219	
chr4	188916669	188916902	*ZFP42*	234	
chr12	7244833	7245222	*CR1*	390	Božić *et al.*, 2015
chr10	102760670	102761270	*LZTS2*	601	
chr9	127531036	127532010	*NR6A1*	975	
chr2	11887332	11888329	*LPIN1*	998	Qu *et al.*, 2017
chr10	102756826	102762363	*LZTS2*	5538	
chr9	127531279	127534048	*NR6A1*	2770	
chr5	1008954	1009718	*NKD2*	765	Li *et al.*, 2017
chr17	48049597	48050018	*DLX4*	422	Zhou *et al.*, 2016
chr2	25499904	25500262	*DNMT3A*	359	Jost *et al.*, 2014
chr19	33794630	33794925	*CEBPA*	296	Lin *et al.*, 2011
chr5	78407502	78407733	*BHMT*	232	Deneberg *et al.*, 2011
chr3	126243341	126243461	*CHST13*	121	
chr3	126242430	126242550	*CHST13*	121	
chr6	56112241	56112534	*COL21A1*	294	
chr20	23433299	23433553	*CST11*	255	
chr5	54281286	54281487	*ESM1*	202	
chr1	204121865	204121985	*ETNK2*	121	
chr1	204120529	204120649	*ETNK2*	121	
chr11	65816582	65816702	*GAL3ST3*	121	
chr12	5020142	5020262	*KCNA1*	121	
chr12	5020633	5020753	*KCNA1*	121	
chr12	4918109	4918229	*KCNA6*	121	
chr12	4918331	4918451	*KCNA6*	121	
chr11	64058818	64058938	*KCNK4*	121	
chr11	64057747	64057867	*KCNK4*	121	
chr5	126626447	126626769	*MEGF10*	323	
chr8	101661418	101661538	*MGC39715*	121	
chr8	101661993	101662113	*MGC39715*	121	
chr5	134871106	134871226	*NEUROG1*	121	
chr5	134871402	134871522	*NEUROG1*	121	
chr5	134871598	134871718	*NEUROG1*	121	
chr5	134870454	134870574	*NEUROG1*	121	
chr5	134870917	134871037	*NEUROG1*	121	
chr5	134872064	134872184	*NEUROG1*	121	

chr13	97646281	97646401	*OXGR1*	121	
chr13	97646751	97646871	*OXGR1*	121	
chr20	3218085	3218205	*SLC4A11*	121	
chr20	3218440	3218560	*SLC4A11*	121	
chr5	121647798	121647918	*SNCAIP*	121	
chr5	121647248	121647368	*SNCAIP*	121	
chr20	591340	591460	*TCF15*	121	
chr20	590183	590303	*TCF15*	121	
chr3	133465409	133465529	*TF*	121	
chr3	133465120	133465240	*TF*	121	
chr8	21881434	21883388	*NPM2*	1955	Kelly *et al.*, 2017
chr5	180016802	180019096	*SCGB3A1*	2295	
chr9	22007977	22009873	*CDKN2B*	1897	
chr9	22004658	22006684	*CDKN2B*	2027	
chr1	36915167	36916986	*OSCP1*	1820	
chr11	14993444	14994500	*CALCA*	1057	Hájková *et al.*, 2012
chr16	11348266	11350810	*SOCS1*	2545	
chr16	68770870	68772866	*CDH1*	1997	
chr11	17740807	17743777	*MYOD1*	2971	
chr6	152128341	152129813	*ESR1*	1473	
chr9	90112514	90114067	*DAPK1*	1554	
chr19	10380285	10381990	*ICAM1*	1706	
chr5	1293465	1295472	*TERT*	2008	
chr5	1282392	1283291	*TERT*	900	
chr22	33196800	33198315	*TIMP3*	1516	
chr5	138088422	138089841	*CTNNA1*	1420	
chr5	137799044	137802235	*EGR1*	3192	
chr11	47414710	47417876	*PU1* enhancer	3167	Curik *et al.*, 2012
chr11	47399455	47400777	*PU1* promoter	1323	
chr5	135415044	135417216	*VTRNA2-1*	2173	Treppendahl *et al.*, 2012
chr3	50373957	50379097	*RASSF1*	5141	Liu *et al.*, 2017
chr8	41165479	41167704	*SFRP1*	2226	Guo *et al.*, 2017
chr4	154709378	154714360	*SFRP2*	4983	
chr5	150395525	150403475	*GPX3*	7951	Zhou *et al.*, 2017
chr14	25142240	25143442	*GZMB* intergenic region	1203	Šestáková *et al.*, 2019
chr9	128509575	128511617	*PBX3*	2043	Hájková *et al.*, 2014
chr11	32440684	32476289	*WT1*	35606	Genes typically associated with AML pathogenesis (*HOX* genes, *WT1*) and genes and regions investigated in our previous research
chr12	133484657	133485928	*CHFR* enhancer	1272	
chr12	133463633	133464858	*CHFR* promoter	1226	
chr9	23820650	23822189	*ELAVL2*	1540	
chr17	56356348	56359520	*MPO*	3173	
chr20	6747931	6749579	*BMP2*	1649	
chr10	90750227	90751818	*FAS*	1592	
chr7	150777267	150778971	*FASTK*	1705	

chr1	92952101	92952781	*GFI1*	681
chr7	27169323	27171201	*HOXA4*	1879
chr7	27172734	27177485	*HOXA4-HOXA5*	4752
chr7	27178290	27181777	*HOXA5*	3488
chr7	27182505	27186030	*HOXA5*	3526
chr7	27154915	27166717	*HOXA3*	11803
chr7	27186376	27192392	*HOXA6*	6017
chr7	27193350	27196990	*HOXA7*	3641
chr7	27197910	27201014	*HOXA7*	3105
chr7	27203696	27209712	*HOXA9*	6017
chr7	27212509	27220211	*HOXA10*	7703
chr7	27223507	27233202	*HOXA11*	9696
chr7	27238643	27261060	*HOXA13*	22418
chr7	27139958	27153767	*HOXA2-A3*	13810
chr7	27134068	27137303	*HOXA1*	3236
chr17	46619215	46632388	*HOXB2-B3*	13174
chr17	46604015	46608713	*HOXB1*	4699
chr17	46640126	46643166	*HOXB3*	3041
chr17	46651549	46660946	*HOXB4*	9398
chr17	46669329	46675685	*HOXB5*	6357
chr17	46679462	46683607	*HOXB6*	4146
chr17	46684713	46698439	*HOXB7-B8*	13727
chr17	46698623	46711520	*HOXB9*	12898
chr17	46719166	46725338	*HOXB* enhancer	6173
chr17	46795719	46806589	*HOXB13*	10871
chr12	54444758	54448668	*HOXC4*	3911
chr12	54440499	54441686	*HOXC4* enhancer	1188
chr12	54409989	54428840	*HOXC5-C6*	18852
chr12	54399726	54403566	*HOXC8*	3841
chr12	54387508	54394839	*HOXC9*	7332
chr12	54378222	54380526	*HOXC10*	2305
chr12	54366423	54369355	*HOXC11*	2933
chr12	54343383	54349527	*HOXC12*	6145
chr12	54332422	54333888	*HOXC13*	1467
chr12	54319785	54322997	*HOXC* enhancer	3213
chr2	177019560	177030824	*HOXD3*	11265
chr2	177042088	177044146	*HOXD-AS*	2059
chr2	177051619	177055302	*HOXD1*	3684
chr2	177011979	177018152	*HOXD4*	6174
chr2	176999849	177005805	*HOXD-AS2*	5957
chr2	176992050	176996166	*HOXD8*	4117
chr2	176985985	176988584	*HOXD9*	2600
chr2	176979270	176982519	*HOXD10*	3250

chr2	176961724	176972338	*HOXD11-D12*	10615
chr2	176943637	176958583	*HOXD13*	14947
chr2	176929340	176937139	*HOXD* enhancer	7800
chr2	85810361	85813304	*VAMP5*	2944
chr1	64058191	64060458	*PGM1*	2268
chr19	55572760	55581129	*RDH13*	8370
chr2	64066479	64070415	*UGP2*	3937
chr10	17268605	17273715	*VIM*	5111
chr12	66581289	66584587	*IRAK3*	3299
chr15	40599247	40601307	*PLCB2* promoter	2061
chr15	40614106	40617359	*PLCB2* enhancer	3254
chr15	40583086	40583845	*PLCB2*	760
chr6	32819966	32823082	*PSMB9-TAP1*	3117
chr16	53467492	53471613	*RBL2*	4122
chr10	73846564	73850018	*SPOCK2*	3455
chr10	73855698	73859352	*SPOCK2* enhancer	3655
chr19	56163319	56168083	*U2AF2*	4765
chr7	148578438	148583983	*EZH2*	5546
chr7	148635825	148640261	*EZH2* enhancer	4437
chr12	56410763	56416635	*IKZF4*	5873
chr8	22018546	22028232	*BMP1*	9687
chr2	66652803	66673827	*MEIS1*	21025
chr22	33196540	33200798	*TIMP3*	4259
chr6	150262167	150263834	*ULBP2*	1668

10 APPENDICES

10.1 Appendix 1

Šestáková Š, Krejčík Z, Folta A, Cerovská E, Šálek C, Merkerová MD et al. **DNA methylation and hydroxymethylation patterns in acute myeloid leukemia patients with mutations in DNMT3A and IDH1/2 and their combinations.** *Cancer biomarkers: section A of Disease markers* **2019**;25(1):43-51.
IF = 3.436 (2019)

https://www.ncbi.nlm.nih.gov/pubmed/30988238

10.2 Appendix 2

Šestáková Š, Šálek C, Remešová H. **DNA Methylation Validation Methods: a Coherent Review with Practical Comparison**. *Biological Procedures Online* **2019**;21(1).
IF = 2.711 (2019)

https://www.ncbi.nlm.nih.gov/pubmed/31582911